

A Lower Bound on the Probability of Error of Polar Codes over BMS Channels

Boaz Shuval, Ido Tal

Department of Electrical Engineering,
Technion, Haifa 32000, Israel.

Email: {bshuval@campus, idotal@ee}.technion.ac.il

Abstract—Polar codes are a family of capacity-achieving codes that have explicit and low-complexity construction, encoding, and decoding algorithms. Decoding of polar codes is based on the successive-cancellation decoder, which decodes in a bit-wise manner. A decoding error occurs when at least one bit is erroneously decoded. The various codeword bits are correlated, yet performance analysis of polar codes ignores this dependence: the upper bound is based on the union bound, and the lower bound is based on the worst-performing bit.

Improvement of the lower bound is afforded by considering error probabilities of two bits simultaneously. These are difficult to compute explicitly due to the large alphabet size inherent to polar codes. In this research we propose a method to lower-bound the error probabilities of bit pairs. We develop several transformations on pairs of synthetic channels that make the resultant synthetic channels amenable to alphabet reduction. Our method improves upon currently known lower bounds for polar codes under successive-cancellation decoding.

Index Terms—Channel polarization, channel upgrading, lower bounds, polar codes, probability of error.

I. INTRODUCTION

POLAR codes [1] are a family of codes that achieve capacity on binary, memoryless, symmetric (BMS) channels and have low-complexity construction, encoding, and decoding algorithms. This is the setting we consider. Polar codes have since been extended to a variety of settings including source-coding [2], [3], non-binary channels [4], asymmetric channels [5], channels with memory [6], [7], and more. The probability of error of polar codes is given by a union of correlated error events. The union bound, which ignores this correlation, is used to upper-bound the error probability. In this work, we exploit the correlation between error events to develop a general method for lower-bounding the probability of error of polar codes.

Polar codes are based on an iterative construction that transforms $N = 2^n$ identical and independent channel uses into “good” and “bad” channels. The “good” channels are almost noiseless, whereas the “bad” channels are almost pure noise. Arkan showed [1] that for every $\epsilon > 0$, as $N \rightarrow \infty$ the proportion of channels with capacity greater than $1 - \epsilon$ tends to the channel capacity C and the proportion of channels with capacity less than ϵ tends to $1 - C$.

The polar construction begins with two identical and independent copies of a BMS W and transforms them into two new channels,

$$\begin{aligned} W^-(y_1, y_2|u_1) &= \frac{1}{2} \sum_{u_2} W(y_1|u_1 \oplus u_2) W(y_2|u_2), \\ W^+(y_1, y_2, u_1|u_2) &= \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2). \end{aligned} \quad (1)$$

Channel W^+ is a better channel than W whereas channel W^- is worse than W .¹ This construction can be repeated multiple times; each time we take two identical copies of a channel, say W^+ and W^+ , and polarize them, e.g., to W^{+-} and W^{++} . We call the operation $W \mapsto W^-$ a ‘-’-transform, and the operation $W \mapsto W^+$ a ‘+’-transform.

There are $N = 2^n$ possible combinations of n ‘-’- and ‘+’-transforms; we define channel W_i as follows. Let $b_1 b_2 \dots b_n$ be the binary expansion of $i - 1$, where b_1 is the most significant bit (MSB). Then, channel W_i is obtained by n transforms of W according to the sequence b_1, b_2, \dots, b_n , starting with the MSB: if $b_j = 0$ we do a ‘-’-transform and if $b_j = 1$ we do a ‘+’-transform. For example, if $n = 3$, channel W_5 is W^{+-} , i.e., it first undergoes a ‘+’-transform and then two ‘-’-transforms.

Overall, we obtain N channels W_1, \dots, W_N ; channel W_i has input u_i and output $y_1, \dots, y_N, u_1, \dots, u_{i-1}$. I.e., channel W_i has binary input u_i , output that consists of the output and input of channel W_{i-1} , and assumes that the input bits of future channels u_{i+1}, \dots, u_N are uniform. We call these *synthetic channels*. One then determines which synthetic channels are “good” and which are “bad”, and transmits information over the “good” synthetic channels and predetermined values over the “bad” synthetic channels. Since the values transmitted over the latter are predetermined, we call the “bad” synthetic channels *frozen*.

Decoding is accomplished via the successive-cancellation (SC) decoder. It decodes the synthetic channels in succession, using previous bit decisions as part of the output. The bit decision for a synthetic channel is either based on its likelihood or, if it is frozen, on its predetermined value. I.e., denoting the set of non-frozen synthetic channels by \mathcal{A} ,

$$\hat{U}_i(y_1^N, \hat{u}_1^{i-1}) = \begin{cases} \arg \max_{u_i} W_i(y_1^N, \hat{u}_1^{i-1}|u_i) & i \in \mathcal{A} \\ u_i & i \in \mathcal{A}^c, \end{cases}$$

An abbreviated version of this article, with the proofs omitted, was submitted to ISIT 2017.

¹By this we mean that channel W^+ can be stochastically degraded to channel W , which in turn can be stochastically degraded to W^- .

where we denoted $y_1^N = y_1, \dots, y_N$ and similarly for the previous bit decisions \hat{u}_1^{i-1} . As non-frozen synthetic channels are almost noiseless, previous bit decisions are assumed to be correct. Thus, when N is sufficiently large, this scheme can be shown to achieve capacity [1] as the proportion of almost noiseless channels is C .

To analyze the performance of polar codes, let \mathcal{B}_i denote the event that channel W_i errs under SC decoding while channels $1, 2, \dots, i-1$ do not. I.e.,

$$\mathcal{B}_i = \left\{ u_1^N, y_1^N \mid \hat{u}_1^{i-1} = u_1^{i-1}, \hat{U}_i(y_1^N, \hat{u}_1^{i-1}) \neq u_i \right\}.$$

The probability of error of polar codes under SC decoding is given by $\mathbb{P}\{\bigcup_{i \in \mathcal{A}} \mathcal{B}_i\}$. Let \mathcal{E}_i denote the event that channel W_i errs given that a genie had revealed to it the true previous bits, i.e.

$$\mathcal{E}_i = \left\{ u_1^N, y_1^N \mid \hat{U}_i(y_1^N, u_1^{i-1}) \neq u_i \right\}.$$

We call an SC decoder with access to genie-provided previous bits a *genie-aided decoder*. Some thought reveals that $\bigcup_{i \in \mathcal{A}} \mathcal{B}_i = \bigcup_{i \in \mathcal{A}} \mathcal{E}_i$ (see [4, Proposition 2.1] or [8, Lemma 1]). Thus, the probability of error of polar codes under SC decoding is equivalently given by $P_e^{\text{SC}}(W) = \mathbb{P}\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\}$. In the sequel we assume a genie-aided decoder.

The events $\{\mathcal{B}_i\}$ are disjoint but difficult to analyze. The events \mathcal{E}_i are easier to analyze, but are no longer disjoint. A straightforward upper bound for $\mathbb{P}\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\}$ is the union bound:

$$\mathbb{P}\left\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\right\} \leq \sum_{i \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_i\}.$$

This bound facilitated the analysis of [1]. An important question is how tight this upper bound is. To this end, one approach is to develop a lower bound to $\mathbb{P}\{\bigcup \mathcal{E}_i\}$, which is what we pursue in this work.

A trivial lower bound on a union is

$$\mathbb{P}\left\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\right\} \geq \max_{i \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_i\}. \quad (2)$$

Better lower bounds may be obtained by considering pairs of error events:

$$\mathbb{P}\left\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\right\} \geq \max_{i, j \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}.$$

Via the inclusion-exclusion principle, one can combine lower bounds on multiple pairs of error events to obtain a better lower bound [9]

$$\mathbb{P}\left\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\right\} \geq \sum_{i \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_i\} - \sum_{\substack{i, j \in \mathcal{A}, \\ i < j}} \mathbb{P}\{\mathcal{E}_i \cap \mathcal{E}_j\}. \quad (3)$$

This can also be cast in terms of unions of error events using $\mathbb{P}\{\mathcal{E}_i \cap \mathcal{E}_j\} = \mathbb{P}\{\mathcal{E}_i\} + \mathbb{P}\{\mathcal{E}_j\} - \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}$.

To our knowledge, to date there have been two attempts to compute a lower bound on the performance of the SC decoder, both based on (3). The first attempt was in [8], using a density evolution approach, and the second attempt in [10] applies

only to the BEC. We briefly introduce these below, but first we explain where the difficulty lies.

The probability $\mathbb{P}\{\mathcal{E}_i\}$ is given by an appropriate functional of the probability distribution of synthetic channel W_i . However, the output alphabet of W_i is very large. If the output alphabet of W is \mathcal{Y} then the output alphabet of W_i has size $|\mathcal{Y}|^N 2^{i-1}$. This quickly grows unwieldy, recalling that $N = 2^n$. It is infeasible to store this probability distribution and it must be approximated. Such approximations are the subject of [11]; they enable one to compute upper and lower bounds on various functionals of the synthetic channel W_i .

To compute probabilities of unions of events, one must know the joint distribution of two synthetic channels. The size of the joint distribution's output alphabet is the product of each synthetic channel's alphabet size, rendering the joint distribution infeasible to store.

The authors of [8] suggested to approximate the joint distribution of pairs of synthetic channels using a density evolution approach. This provides an iterative method to compute the joint distribution, but does not address the problem of the amount of memory required to store it. Practical implementation of density evolution must involve quantization [12, Appendix B]. The probability of error derived from quantized joint distributions approximates, but does not generally bound, the real probability of error. For the special case of the BEC, as noted and analyzed in [8], no quantization is needed, as the polar transform of a BEC is a BEC. Thus, they were able to precisely compute the probabilities of unions of error events of descendants of a BEC using density evolution.

The same bounds for the BEC were developed in [10] using a different approach, again relying on the property that the polar transform of a BEC is a BEC. The authors were able to track the joint probability of erasure during the polarization process. Furthermore, they were able to show that the union bound is asymptotically tight for the BEC.

In this work, we develop an algorithm to compute lower bounds on the joint probability of error of two synthetic channels $\mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}$. Our technique is general, and applies to synthetic channels that are polar descendants of any BMS channel. We use these bounds in (3) to lower-bound the probability of error of polar codes. For the special case of the BEC, we recover the results of [8] and [10] using our bounds.

Our method is based on approximating the joint distribution with a stochastically upgraded joint distribution that has a smaller output alphabet. The difficulty is that key ideas that are true for single channels no longer apply to joint distributions. For example, a degrading operation on a joint distribution may *improve* the performance of an SC decoder. As another example, a sufficient statistic for a single synthetic channel is not a sufficient statistic for the joint distribution: two symbols that are indistinguishable for one synthetic channel may have very different meanings for future synthetic channels. Therefore, we develop methods that in one sense decouple the two synthetic channels yet in another sense couple them even further.

II. OVERVIEW OF OUR METHOD

In this section we provide a brief overview of our method, and lay out the groundwork for the sections that follow. We aim to produce a lower bound on the probability of error of two synthetic channels. Since we cannot know the precise joint distribution, we must approximate it. The approximation is rooted in stochastic degradation.

Degradation is a partial ordering of channels. Let $W(y|u)$ and $Q(z|u)$ be two channels. We say that W is (stochastically) degraded with respect to Q , denoted $W \preceq Q$, when there exists some channel $P(y|z)$ such that

$$W(y|u) = \sum_z P(y|z)Q(z|u). \quad (4)$$

If W is degraded with respect to Q then Q is upgraded with respect to W . Degradation implies an ordering on the probability of error of the channels [12, Chapter 4]: if $W \preceq Q$ then $P_e(W) \geq P_e(Q)$. This is true only when the decoder used is the optimal decoder.

The notion of degradation readily extends to joint channels. We say that joint channel $Q_{a,b}(z_a, z_b|u_a, u_b) \succcurlyeq W_{a,b}(y_a, y_b|u_a, u_b)$ via some degrading channel $P(y_a, y_b|z_a, z_b)$ if

$$W_{a,b}(y_a, y_b|u_a, u_b) = \sum_{z_a, z_b} P(y_a, y_b|z_a, z_b)Q_{a,b}(z_a, z_b|u_a, u_b). \quad (5)$$

As for the single channel case, if $Q_{a,b} \succcurlyeq W_{a,b}$ then $P_e(W_{a,b}) \geq P_e(Q_{a,b})$, where P_e is the probability of error of the optimal decoder for the joint channel. Indeed our approach will be to approximate the joint synthetic channel with an upgraded joint channel with smaller output alphabet. There is a snag, however: this ordering of error probabilities does not hold, in general, for suboptimal decoders.

The SC decoder, used for polar codes, is suboptimal. In the genie-aided case, which we consider here, it is equivalent to performing a maximum likelihood decision on each marginal separately. We shall demonstrate the suboptimality of the SC decoder in Section III. Then, we will develop a different decoding criterion whose performance lower-bounds the SC decoder performance and is ordered by degradation. While in general this decoder requires an exhaustive search, for the special case of polar codes this decoder is easily found. It does, however, imply a special structure for the degrading channel, which we use to our advantage.

We investigate the joint distribution of two synthetic channels in Section IV. We first bring it to a more convenient form that will be used in the sequel. Then, we explain how to polarize a joint synthetic channel distribution and explore some consequences of symmetry. Further consequences of symmetry are the subject of Section V, in which we transform the channel to another form that greatly simplifies the steps that follow. This form exposes the inherent structure of the joint distribution.

How to actually upgrade joint distributions is the subject of Section VI. We upgrade the joint distribution in two ways; each upgrades one marginal without changing the other. We cannot simply upgrade the marginals, as we must consider the joint distribution as a whole. This is where the above-mentioned methods of coupling-decoupling come into play.

We present our algorithm for lower-bounding the probability of error of polar codes in Section VII. This algorithm is based on the building blocks presented in the previous sections. We demonstrate our algorithm with some numerical results in Section VIII.

A. Notation

We denote by $y_j^k = y_j, y_{j+1}, \dots, y_k$ for $j < k$. We use an Iverson-style notation (see [13]) for indicator (characteristic) functions. I.e., for a logical expression expr , $\llbracket \text{expr} \rrbracket$ is 0 whenever expr is not true and is 1 otherwise. We assume that the indicator function takes precedence whenever it appears, e.g., $n^{-1} \llbracket n > 0 \rrbracket$ is 0 for $n = 0$.

III. DECODING OF TWO DEPENDENT CHANNELS

In this section, we tackle decoding of two dependent channels. We explain how this differs from the case of decoding a single channel, and dispel some misconceptions that may arise. We then specialize the discussion to polar codes. We explain the difficulty with combining the SC decoder with degradation procedures, and develop a different decoding criterion instead. Finally, we develop a special structure for the degrading channel that, combined with the decoding criterion, implies ordering of probability of error by degradation.

A. General Case

A decoder for channel $W : \mathcal{U} \rightarrow \mathcal{Y}$ is a mapping ϕ that maps every output sequence $y \in \mathcal{Y}$ to some $u \in \mathcal{U}$. The average probability of error of the decoder for equiprobable inputs is given by

$$P_e = \frac{1}{|\mathcal{U}|} \sum_u \sum_y W(y|u) \mathbb{P} \{ \phi(y) \neq u \}.$$

The decoder is deterministic for symbols y where $\mathbb{P} \{ \phi(y) \neq u \}$ assumes only the values 0 and 1. For some symbols, however, we allow the decoder to make a random decision. If $W(y|u) = W(y|u')$ for some $u, u' \in \mathcal{U}$, then P_e is the same whether $\phi(y) = u$ or $\phi(y) = u'$. Thus, the probability of error is insensitive to the resolution of ties. We denote the error event of a decoder by $\mathcal{E} = \{(u, y) : \phi(y) \neq u\}$. It is dependent on the decoder, i.e., $\mathcal{E} = \mathcal{E}(\phi)$; we suppress this to avoid cumbersome notation. Clearly, $P_e = \mathbb{P} \{ \mathcal{E} \}$.

The maximum-likelihood (ML) decoder, well known to minimize P_e when the input bits are equiprobable, is defined by

$$W(y|u) > W(y|u') \quad \forall u' \neq u \Rightarrow \phi(y) = u. \quad (6)$$

The ML decoder is not unique, as it does not define how ties are resolved.

We now consider two *dependent* binary-input channels, $W_a : \mathcal{U} \rightarrow \mathcal{Y}_a$ and $W_b : \mathcal{U} \rightarrow \mathcal{Y}_b$, with joint distribution $W_{a,b} : \mathcal{U} \times \mathcal{U} \rightarrow \mathcal{Y}_a \times \mathcal{Y}_b$. The optimal decoder for the joint channel considers both outputs together and makes a decision for both inputs jointly; its probability of error is $P_e(W_{a,b})$. Rather than jointly decoding the input bits based on the joint output, we may opt to decode each channel separately. I.e., the decoder of channel W_a bases its decision solely on y_a and completely

ignores y_b and vice versa. What are the optimal decoders ϕ_a and ϕ_b ? The answer depends on the criterion of optimality.

Denote by \mathcal{E}_i the error event of channel W_i under some decoder $\phi_i : \mathcal{Y}_i \rightarrow \mathcal{U}$. If we wish to minimize each individual channel's probability of error, we set each decoder as the ML decoder for the respective channel. We call this the *Individual Maximum Likelihood* (IML) decoder, and denote its probability of error by $P_e^{\text{IML}}(W_{a,b})$. Another criterion is to minimize $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$, the probability that at least one of the decoders makes an error. We call the decoder that minimizes this probability using individual decoders for each channel the *Individual Minimum Joint* P_e (IMJP) decoder. The event $\mathcal{E}_a \cup \mathcal{E}_b$ is not the same as the error event of the optimal decoder for the joint channel, even when the individual decoders turn out to be ML decoders. This is because we decode each input bit separately using only a portion of the joint output. Clearly,

$$P_e(W_{a,b}) \leq \min_{\phi_a, \phi_b} \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\} \leq P_e^{\text{IML}}(W_{a,b}). \quad (7)$$

The three decoders in (7) successively use less information for their decisions. The optimal decoder uses both outputs jointly as well as knowledge of the joint probability distribution; the IMJP decoder retains the knowledge of the joint distribution, but uses each output separately; finally, the IML decoder dispenses with the joint distribution and operates as if the marginals are independent channels.

Example 1. The conditional distribution $W_{a,b}(y_a, y_b|u_a, u_b)$ of some joint channel is given in Table I.² The marginals are channels $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$. The optimal decoder for the joint channel chooses, for each output pair, the input pair with the highest probability; it achieves $P_e(W_{a,b}) = 0.52$. It is easily verified that the ML decoders of the marginals decide that the input is 0 when 1 is received and vice versa; thus, $P_e^{\text{IML}}(W_{a,b}) = 0.7075$. If we change the decoder of channel W_b to always declare 0, regardless of the output, then $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\} = 0.6575$. By checking all combinations of decoders ϕ_a, ϕ_b , it can be verified that this is indeed the minimum value of $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$. As expected, (7) holds.

We now demonstrate that the probability of error of suboptimal decoders is not ordered by degradation. To this end, we degrade the joint channel in Table I by merging the output symbols (0,0), (1,1) into a new symbol, (0',0') and (0,1), (1,0) into a new symbol, (1',1'). Denote the new joint channel $W'_{a,b}$. For each of the marginals, the ML decoder declares 0 upon receipt of 0', and 1 otherwise. Hence, for the degraded channel, $P_e^{\text{IML}}(W'_{a,b}) = 0.555$, which is *lower* than $P_e^{\text{IML}}(W_{a,b})$. For the degraded channel, the IML decoder is also the optimal decoder. As this is a degraded channel, however, $P_e^{\text{IML}}(W'_{a,b}) = P_e(W'_{a,b}) \geq P_e(W_{a,b}) = 0.52$.

B. Polar Coding Setting

Given a joint channel distribution, finding an optimal or IML decoder is an easy task. In both cases we use maximum-likelihood decoders; in the first case based on the joint

TABLE I
CONDITIONAL DISTRIBUTION $W_{a,b}(y_a, y_b|u_a, u_b)$. IN THIS CASE, THE ML DECODERS OF THE MARGINALS DO NOT MINIMIZE $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$.

| (u_a, u_b) | (y_a, y_b) | | | |
|--------------|--------------|--------|--------|--------|
| | (0, 0) | (0, 1) | (1, 0) | (1, 1) |
| (0, 0) | 0.30 | 0.04 | 0.04 | 0.62 |
| (0, 1) | 0.44 | 0.46 | 0.01 | 0.09 |
| (1, 0) | 0.22 | 0.49 | 0.24 | 0.05 |
| (1, 1) | 0.05 | 0.54 | 0.32 | 0.09 |

distribution, whereas in the second case based on the marginal distributions. On the other hand, finding an IMJP decoder requires an exhaustive search, which may be costly. In the polar coding setting, as we now show, the special structure of joint synthetic channels permits finding the IMJP decoder without resorting to a search procedure.

1) *Joint Distribution of Two Synthetic Channels:* Let W be some BMS channel that undergoes n polarization steps. Let a and b be two indices of synthetic channels, where $b > a$. The synthetic channels are $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$, where $y_a = (y_1^N, u_1^{a-1})$, $y_b = (y_1^N, u_1^{b-1})$, and $N = 2^n$. We call them the *a-channel* and the *b-channel*, respectively. Their joint distribution is $W_{a,b}(y_a, y_b|u_a, u_b)$. I.e., this is the probability that the output of the a-channel is y_a and the output of the b-channel is y_b , given that the inputs to the channels are u_a and u_b , respectively.

With probability 1, the prefix of y_b is (y_a, u_a) . Namely, y_b has the form

$$y_b = ((y_1^N, u_1^{a-1}), u_a, u_{a+1}^{b-1}) \equiv (y_a, u_a, y_r),$$

where y_r denotes the remainder of y_b after removing y_a and u_a . Thus,

$$W_{a,b}(y_a, y_b|u_a, u_b) = 2W_b(y_b|u_b) \mathbb{I}[y_b = (y_a, u_a, y_r)], \quad (8)$$

for some arbitrary y_r . The factor 2 stems from the uniform distribution of u_a . With some abuse of notation, we will write

$$\begin{aligned} W_{a,b}(y_a, y_b|u_a, u_b) &= W_{a,b}(y_b|u_a, u_b) \\ &= W_{a,b}(y_a, u_a, y_r|u_a, u_b). \end{aligned}$$

The right-most expression makes it clear that the portion of y_b in which the input of the a-channel appears must equal the actual input of the a-channel.

Observe from (8) that we can think of $W_b(y_a, u_a, y_r|u_b)$ as the joint distribution $W_{a,b}$ up to a constant factor. Indeed, we will use $W_b(y_a, u_a, y_r|u_b)$ to denote the joint channel where convenient.

2) *Decoders for Joint Synthetic Channels:* Which decoders can we consider for joint synthetic channels? The optimal decoder extracts u_a from the output of the b-channel and proceeds to decode u_b . This outperforms the SC decoder but is also impractical and does not lend itself to computing the probability that is of interest to us, the probability that *either* of the synthetic channels errs. A natural suggestion is to mimic the SC decoder, i.e., to use an IML decoder. The joint probability of error of this decoder may decrease after stochastic degradation, so we discard this option.

²This is not a joint distribution of two synthetic channels that result from polarization. However, the phenomena observed here hold for joint distributions of two synthetic channels as well, and similar examples may be constructed for the polar case.

Consider two decoders ϕ_a and ϕ_b for channels W_a and W_b , respectively. As above, \mathcal{E}_i is the error event of channel W_i using decoder ϕ_i . We seek a lower bound on $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$. Therefore, we choose decoders ϕ_a and ϕ_b that minimize $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$; this is none other than the IMJP decoder. Its performance lower-bounds that of the IML decoder (see (7)). As we shall later see, combined with a suitable degrading channel structure, the probability of error of the IMJP decoder increases after stochastic degradation. Conversely, it decreases under stochastic upgradation; thus, combining the IMJP decoder with a suitable upgrading procedure produces the desired lower bound.

Multiple decoders may achieve $\min_{\phi_a, \phi_b} \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$. One decoder can be found in a straight-forward manner; we call it *the* IMJP decoder. The following theorem shows how to find it. Its proof is a direct consequence of Lemmas 3 and 4 that follow.

Theorem 1. *Let $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ be two channels with joint distribution $W_{a,b}$ that satisfies (8). Then, $\min_{\phi_a, \phi_b} \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$ is achieved by setting ϕ_b as an ML decoder for W_b and ϕ_a according to*

$$\phi_a(y_a) = \arg \max_{u_a} T(y_a|u_a), \quad (9)$$

where

$$T(y_a|u_a) = \frac{1}{2} \sum_{\substack{u_b, \\ y_b}} W_{a,b}(y_a, y_b|u_a, u_b) \mathbb{P}\{\phi_b(y_b) = u_b\}. \quad (10)$$

Note that $T(y_a|u_a)$ is not a conditional distribution; it is non-negative, but its sum over y_a does not necessarily equal 1.

Corollary 2. *Theorem 1 holds for any two synthetic channels $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ that result from the same number of polarization steps of a BMS, where index b is greater than a .*

Proof: In the polar code case, the joint channel satisfies (8), so Theorem 1 applies. ■

In what follows, denote

$$\varphi_i(y_i, u_i) \triangleq \mathbb{P}\{\phi_i(y_i) = u_i\}, \quad i = a, b.$$

Lemma 3. *Let $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ be two dependent binary-input channels with equiprobable inputs and joint distribution $W_{a,b}$ that satisfies (8). Let $\phi_a : \mathcal{Y}_a \rightarrow \mathcal{U}$ be some decoder for channel W_a with error event \mathcal{E}_a . Then, setting ϕ_b as an ML decoder for W_b achieves $\min_{\phi_b} \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$.*

Proof: Recall that $y_b = (y_a, u_a, y_r)$. Using (8),

$$\begin{aligned} 1 - \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\} &= \frac{1}{4} \sum_{\substack{u_a, \\ u_b}} \sum_{\substack{y_a, \\ y_b}} W_{a,b}(y_a, y_b|u_a, u_b) \varphi_a(y_a, u_a) \varphi_b(y_b, u_b) \\ &= \frac{1}{2} \sum_{\substack{u_a, \\ y_a, y_b}} \varphi_a(y_a, u_a) \mathbb{P}\{y_b = (y_a, u_a, y_r)\} g(y_b), \end{aligned}$$

where

$$g(y_b) = \sum_{u_b} \varphi_b(y_b, u_b) W_b(y_b|u_b).$$

The problem of finding the decoder ϕ_b that minimizes $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$ is separable over u_a, y_a, y_b ; the terms $\varphi_a(y_a, u_a)$, $\mathbb{P}\{y_b = (y_a, u_a, y_r)\}$ are non-negative and independent of u_b . Therefore, the optimal decoder ϕ_b is given by $\phi_b(y_b) = \arg \max_{u'_b} W_b(y_b|u'_b)$. ■

Lemma 4. *Let $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ be two binary-input channels with joint distribution $W_{a,b}(y_a, y_b|u_a, u_b)$ and equiprobable inputs. Let $\phi_b : \mathcal{Y}_b \rightarrow \mathcal{U}$ be some decoder for channel W_b . Then, the decoder ϕ_a for channel W_a given by (9) minimizes $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$.*

Proof: Since the input is equiprobable,

$$\begin{aligned} 1 - \mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\} &= \frac{1}{4} \sum_{\substack{u_a, \\ y_a}} \sum_{\substack{u_b, \\ y_b}} W_{a,b}(y_a, y_b|u_a, u_b) \varphi_a(y_a, u_a) \varphi_b(y_b, u_b) \\ &= \frac{1}{2} \sum_{\substack{u_a, \\ y_a}} \varphi_a(y_a, u_a) \cdot \frac{1}{2} \sum_{\substack{u_b, \\ y_b}} W_{a,b}(y_a, y_b|u_a, u_b) \varphi_b(y_b, u_b) \\ &= \frac{1}{2} \sum_{\substack{u_a, \\ y_a}} T(y_a|u_a) \varphi_a(y_a, u_a), \end{aligned}$$

where the last equality is by (10). The problem of finding the decoder ϕ_a that minimizes $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$ is separable over y_a ; clearly the optimal decoder is the one that sets $\phi_a(y_a) = \arg \max_{u'_a} T(y_a|u'_a)$. ■

Using (8), if ϕ_b is chosen as an ML decoder, as per Lemma 3, we have the following expression for $T(y_a|u_a)$:

$$\begin{aligned} T(y_a|u_a) &= \sum_{y_r} \sum_{u_b} W_b(y_a, u_a, y_r|u_b) \varphi_b(y_b, u_b) \\ &= \sum_{y_r} \max_{u_b} W_b(y_a, u_a, y_r|u_b). \end{aligned} \quad (11)$$

The IMJP and IML decoders do not coincide in general, although in some cases they may indeed coincide. We demonstrate this in the following example.

Example 2. Let W be a BSC with crossover probability p . We perform $n = 2$ polarization steps and consider the joint channel $W_{1,4}$, i.e. $W_a = W^{--}$ and $W_b = W^{++}$. When $p = 0.4$, we have $0.6544 = P_e^{\text{IMJP}}(W_{1,4}) < P_e^{\text{IML}}(W_{1,4}) = 0.6976$. On the other hand, when $p = 0.2$, $P_e^{\text{IMJP}}(W_{1,4}) = P_e^{\text{IML}}(W_{1,4}) = 0.5136$. In either case, (7) holds.

Remark 1. In the special case where W is a BEC and W_a and W_b are two of its polar descendants, the IMJP and IML (SC) decoders coincide. This is thanks to a special property of the BEC that erasures for a synthetic channel are determined by the outputs of the $N = 2^n$ copies of a BEC, regardless of the inputs of previous synthetic channels. We show this in Appendix A.

3) Proper Degrading Channels: The IMJP decoder is attractive for joint polar synthetic channels since, by Theorem 1, we can efficiently compute it. This was made possible by the successive form of the joint channel (8). Thus, we seek degrading channels that maintain this form.

Let $W_{a,b}(y_a, y_b|u_a, u_b)$ be a joint distribution of two synthetic channels and let $Q_{a,b}(z_a, z_b|u_a, u_b) \succcurlyeq$

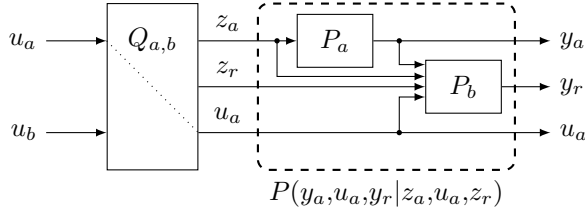


Fig. 1. The structure of proper degrading channels.

$W_{a,b}(y_a, y_b | u_a, u_b)$. The marginal channels of $Q_{a,b}$ are $Q_a(z_a | u_a)$ and $Q_b(z_b | u_b)$. The most general degrading channel is of the form

$$P(y_a, y_b | z_a, z_b) = P_1(y_a | z_a, z_b) \cdot P_2(y_b | z_a, z_b, y_a),$$

where P_1 and P_2 are probability distributions. This form does not preserve the successive structure of joint synthetic channels (8). Even if $Q_{a,b}$ satisfies (8), the resulting $W_{a,b}$ may not. To this end, we turn to a subset of degrading channels. Recalling that $y_b = (y_a, u_a, y_r)$, we consider degrading channels of the form

$$\begin{aligned} P(y_a, u_a, y_r | z_a, u_a, z_r) \\ = P_a(y_a | z_a) \cdot P_b(y_r | z_a, u_a, z_r, y_a). \end{aligned} \quad (12)$$

I.e., these degrading channels degrade z_a , the output of Q_a , to y_a , pass u_a unchanged, and degrade z_r , the remainder of Q_b 's output, to y_r . For this to be a valid channel, P_a and P_b must be probability distributions. This degrading channel structure is illustrated in Figure 1. By construction, degrading channels of the form (12) preserve the form (8) that is required for efficiently computing the IMJP decoder as in Theorem 1.

Definition 1 (Proper degrading channels). A degrading channel of the form (12) is called *proper*. We write $Q \succcurlyeq W$ to denote that channel Q is upgraded from W with a proper degrading channel. We say that an upgrading (degrading) procedure is proper if its degrading channel is proper.

By marginalizing the joint distribution it is straight-forward to deduce the following for joint synthetic channel distributions.

Lemma 5. If joint channel $Q_{a,b}(z_a, u_a, z_r | u_a, u_b) \stackrel{p}{\succcurlyeq} W_{a,b}(y_a, u_a, y_r | u_a, u_b)$, then $Q_a(z_a | u_a) \succcurlyeq W_a(y_a | u_a)$ and $Q_b(z_a, u_a, z_r | u_b) \succcurlyeq W_b(y_a, u_a, y_r | u_b)$.

This lemma is encouraging, but insufficient for our purposes. It is easy to take degrading channels that are used for degrading a single (not joint) synthetic channel and cast them into a proper degrading channel for joint distributions. This, however, is not our goal. Instead, we start with $W_{a,b}$ and seek an *upgraded* $Q_{a,b}$ with smaller output alphabet that can be degraded to $W_{a,b}$ using a proper degrading channel. This is a very different problem than the degrading one, and its solution is not immediately apparent. Plain-vanilla attempts to use upgrading procedures for single channels fail to produce the desired results. Later, we develop proper upgrading procedures that upgrade one of the marginals without changing the other.

We now show that the probability of error of the IMJP decoder does not decrease after degradation by proper degrading

channels. Intuitively, this is because the decoder for the original channel can simulate the degrading channel. We denote by \mathcal{E}_a^W the error event of channel W_a under some decoder ϕ_a , and similarly define \mathcal{E}_a^Q , \mathcal{E}_b^W , and \mathcal{E}_b^Q . Further, we denote by ϕ_i decoders for W_i and by ψ_i decoders for Q_i , $i = a, b$.

Lemma 6. Let joint channel $W_{a,b}(y_a, u_a, y_r | u_a, u_b)$ have marginals $W_a(y_a | u_a)$ and $W_b(y_a, u_a, y_r | u_b)$. Assume that $Q_{a,b}(z_a, u_a, z_r | u_a, u_b) \stackrel{p}{\succcurlyeq} W_{a,b}(y_a, u_a, y_r | u_a, u_b)$, then $\min_{\psi_a, \psi_b} \mathbb{P} \{ \mathcal{E}_a^Q \cup \mathcal{E}_b^Q \} \leq \min_{\phi_a, \phi_b} \mathbb{P} \{ \mathcal{E}_a^W \cup \mathcal{E}_b^W \}$.

Proof: The proof follows by noting that for any decoder ϕ_i , $i = a, b$ we can find a decoder ψ_i with identical performance. First consider the decoder for channel a . Denote by $\arg P_a(y_a | z_a)$ the result of drawing y_a with probability $P_a(\cdot | z_a)$. Then, $\psi_a(z_a) = \phi_a(\arg P_a(y_a | z_a))$; i.e., this is the decoder that results from first degrading the a -channel output and only then decoding. Next, consider the decoder for the b -channel. Denote by $\arg P_b(y_r | z_a, u_a, z_r, y_a)$ the result of drawing y_r with probability $P_b(\cdot | z_a, u_a, z_r, y_a)$. Then, similar to the a -channel case, $\psi_b(z_a, u_a, z_r) = \phi_b(\arg P_a(y_a | z_a), u_a, \arg P_b(y_r | z_a, u_a, z_r, y_a))$. Hence, the best decoder pair ψ_a, ψ_b cannot do worse than the best decoder pair ϕ_a, ϕ_b . ■

Let W be a BMS channel that undergoes n polarization steps. The probability of error of a polar code with non-frozen set \mathcal{A} under SC decoding is given by $P_e^{\text{SC}}(W) = \mathbb{P} \{ \bigcup_{a \in \mathcal{A}} \mathcal{E}_a^{\text{ML}} \}$, where $\mathcal{E}_a^{\text{ML}}$ is the error probability of synthetic channel W_a under ML decoding. Obviously, for any $\mathcal{A}' \subseteq \mathcal{A}$,

$$P_e^{\text{SC}}(W) \geq \mathbb{P} \left\{ \bigcup_{a \in \mathcal{A}'} \mathcal{E}_a^{\text{ML}} \right\}. \quad (13)$$

We have already mentioned the simplest such lower bound, $P_e^{\text{SC}}(W) \geq \max_{a \in \mathcal{A}} \mathbb{P} \{ \mathcal{E}_a^{\text{ML}} \}$. We now show that the IMJP decoder provides a tighter lower bound. To this end, denote $P_e^{\text{IMJP}}(W_{a,b}) = \min_{\phi_a, \phi_b} \mathbb{P} \{ \mathcal{E}_a \cup \mathcal{E}_b \}$, where \mathcal{E}_i is the probability of error of channel i under decoder ϕ_i .

Lemma 7. Let W be a BMS channel that undergoes n polarization steps, and let \mathcal{A} be the set of non-frozen bits. Then,

$$P_e^{\text{SC}}(W) \geq \max_{a,b \in \mathcal{A}} P_e^{\text{IMJP}}(W_{a,b}) \geq \max_{a \in \mathcal{A}} \mathbb{P} \{ \mathcal{E}_a^{\text{ML}} \}. \quad (14)$$

Proof: Using (13), $P_e^{\text{SC}}(W) \geq \max_{a,b \in \mathcal{A}} \mathbb{P} \{ \mathcal{E}_a^{\text{ML}} \cup \mathcal{E}_b^{\text{ML}} \}$. By definition, the IMJP decoder seeks decoders ϕ_a and ϕ_b that minimize the joint probability of error of synthetic channels with indices a and b . Therefore, for any two indices a and b we have $\mathbb{P} \{ \mathcal{E}_a^{\text{ML}} \cup \mathcal{E}_b^{\text{ML}} \} \geq P_e^{\text{IMJP}}(W_{a,b})$. In particular, this holds for the indices a, b that maximize the right-hand-side. This establishes the leftmost inequality of (14).

To establish the rightmost inequality of (14), we first show that for any a, b ,

$$P_e^{\text{IMJP}}(W_{a,b}) \geq \max \{ \mathbb{P} \{ \mathcal{E}_a^{\text{ML}} \}, \mathbb{P} \{ \mathcal{E}_b^{\text{ML}} \} \}. \quad (15)$$

To see this, first recall that the IMJP decoder performs ML decoding on the b -channel, yielding $P_e^{\text{IMJP}}(W_{a,b}) \geq \mathbb{P} \{ \mathcal{E}_b^{\text{ML}} \}$.

Next, we construct $W'_{a,b} \stackrel{p}{\succ} W_{a,b}$ in which the b -channel is noiseless, by augmenting the y_r portion of its with u_b , i.e.,

$$W'_{a,b}(y_a, u_a, (y_r, v_b)|u_a, u_b) \\ = W_{a,b}(y_a, u_a, y_r|u_a, u_b) \llbracket v_b = u_b \rrbracket.$$

Channel $W'_{a,b}$ can be degraded to $W_{a,b}$ using a proper degrading channel by omitting v_b from the y_r portion of the output and leaving y_a unchanged. Thus, $P_e^{\text{IMJP}}(W_{a,b}) \geq P_e^{\text{IMJP}}(W'_{a,b}) = \mathbb{P}\{\mathcal{E}_{a_0}^{\text{ML}}\}$.

Finally, denote $a_0 = \arg \max_{a \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_a^{\text{ML}}\}$. By (15), for any $c > a_0 > d$ we have $P_e^{\text{IMJP}}(W_{a_0,c}) \geq \mathbb{P}\{\mathcal{E}_{a_0}^{\text{ML}}\}$ and $P_e^{\text{IMJP}}(W_{d,a_0}) \geq \mathbb{P}\{\mathcal{E}_{a_0}^{\text{ML}}\}$. Since $\max_{a,b \in \mathcal{A}} P_e^{\text{IMJP}}(W_{a,b}) \geq \max_{c,d} \{P_e^{\text{IMJP}}(W_{a_0,c}), P_e^{\text{IMJP}}(W_{d,a_0})\}$ we obtain the proof. ■

Lemmas 6 and 7 are instrumental for our lower bound, which combines upgrading operations and the IMJP decoder.

IV. PROPERTIES OF JOINT SYNTHETIC CHANNELS

In this section, we study the properties of joint synthetic channels. We begin by bringing the joint synthetic channel into an equivalent form where the b -channel's ML decision is immediately apparent. We then explain how to jointly polarize synthetic channels. Finally, we describe some consequences of symmetry on joint distributions and on the IMJP decoder.

A. Representation of Joint Synthetic Channel Distribution using D -values

Two channels W and W' with the same input alphabet but possibly different output alphabets are called *equivalent* if $W \succ W'$ and $W' \succ W$. We denote this by $W \equiv W'$. Channel equivalence can cast a channel in a more convenient form. For example, if W is a BMS, one can transform it to an equivalent channel whose output is a sufficient statistic, such as a D -value (see Appendix B), in which case the ML decoder's decision is immediately apparent.

Let $W_{a,b}(y_a, u_a, y_r|u_a, u_b)$ be a joint synthetic channel. Since the joint distribution is determined by the distribution of W_b , we can transform $W_{a,b}$ to an equivalent channel in which the b -channel D -value³ of symbol (y_a, u_a, y_r) is immediately apparent.

Definition 2 (D -value representation). Joint channel $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ is in D -value representation if the marginal W_b satisfies

$$d_b = \frac{W_b(y_a, u_a, d_b|0) - W_b(y_a, u_a, d_b|1)}{W_b(y_a, u_a, d_b|0) + W_b(y_a, u_a, d_b|1)}.$$

We use the same notation $W_{a,b}$ for both the regular and the D -value representations of the joint channel due to their equivalence. The discussion of the various representations of joint channels in Section III-B applies here as well. In particular, we will frequently use $W_b(y_a, u_a, d_b|u_b)$ to denote the joint synthetic channel distribution.

The following lemma affords a more convenient description of the joint channel, in which, in line with the IMJP decoder, the

b -channel's ML decision is immediately apparent. Moreover, this description greatly simplifies the expressions that follow.

Lemma 8. Channels $W_{a,b}(y_a, u_a, y_r|u_a, u_b)$ and $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ are equivalent and the degrading channels from one to the other are proper.

Proof: To establish equivalence we show that each channel is degraded from the other using proper degrading channels. The only portion of interest in (12) is P_b , as in either direction y_a and u_a are unchanged by the degrading channel. Denote by $D_{y_a, u_a}^{d_b}$ the set of all symbols y_r such that the b -channel D -value of (y_a, u_a, y_r) is d_b , for fixed y_a, u_a . Then,

$$W_{a,b}(y_a, u_a, d_b|u_a, u_b) \\ = \sum_{D_{y_a, u_a}^{d_b}} W_{a,b}(y_a, u_a, y_r|u_a, u_b) \\ = \sum_{y_r} W_{a,b}(y_a, u_a, y_r|u_a, u_b) \cdot P_b(d_b|y_r, y_a, u_a),$$

where

$$P_b(d_b|y_r, y_a, u_a) = \llbracket y_r \in D_{y_a, u_a}^{d_b} \rrbracket.$$

Clearly, the b -channel D -value of (y_a, u_a, d_b) is d_b .

On the other hand, by (8) and since all symbols in $D_{y_a, u_a}^{d_b}$ share the same b -channel D -value,

$$W_{a,b}(y_a, u_a, y_r|u_a, u_b) \\ = \sum_{d_b} W_{a,b}(y_a, u_a, d_b|u_a, u_b) \cdot P'_b(y_r|d_b, y_a, u_a),$$

where

$$P'_b(y_r|d_b, y_a, u_a) = \frac{W_b(y_a, u_a, y_r)}{\sum_{D_{y_a, u_a}^{d_b}} W_b(y_a, u_a, y_r)} \llbracket y_r \in D_{y_a, u_a}^{d_b} \rrbracket,$$

and $W_b(y_a, u_a, y_r) = \frac{1}{2} \sum_{u_b} W_b(y_a, u_a, y_r|u_b)$. ■

Remark 2. In the sequel, we will use this lemma to convert to D -value representation the result of polarizing a joint distribution in D -value representation (see Section IV-B). This is possible because Lemma 8 holds for any representation of $W_{a,b}(y_a, u_a, y_r|u_a, u_b)$ in which u_a, y_a are the input and output, respectively, of the a -channel, u_b is the input of the b -channel, and (y_a, u_a, y_r) is the output of the b -channel. In particular, y_r need not consist of inputs to channels W_{a+1}, \dots, W_{b-1} .

Remark 3. At this point the reader may wonder why we have stopped here and not converted the a -channel output to its D -value. The reason is that this constitutes a degrading operation, which is the opposite of what we need. Two a -channel symbols with the same a -channel D -value may have very different meanings for the IMJP decoder. Thus, we cannot combine them to a single symbol without incurring loss.

When the joint distribution is in D -value representation, proper degrading channels admit the form

$$P(y_a, u_a, d_b|z_a, u_a, z_b) = P_a(y_a|z_a)P_b(d_b|z_a, y_a, u_a, z_b). \quad (16)$$

It is obvious that all properties obtained from degrading channels of the form (12) are retained for degrading channels of

³By “ b -channel D -value” we mean the D -value computed for channel W_b . Instead of D -values, other sufficient statistics of the b -channel could have been used. Our use of D -values was prompted by their bounded range $[-1, 1]$.

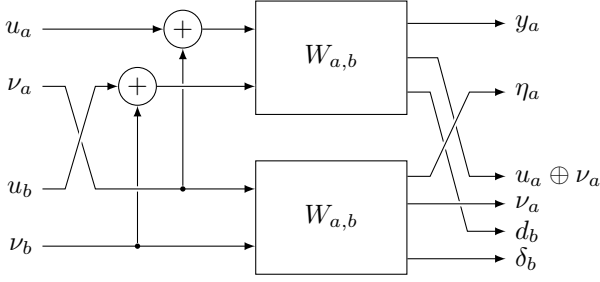


Fig. 2. Polar construction applied jointly to W_a and W_b with joint distribution $W_{a,b}$. The two joint channels are independent duplicates; their inputs are combined using a $(u \oplus v)$ construction.

the form (16). By Lemma 8, we may assume that the degraded channel is also in D -value representation.

B. Polarization for Joint Synthetic Channels

Let $W_{a,b}(y_a, u_a, d_b | u_a, u_b)$ be some joint synthetic channel distribution in D -value representation. We wish to find the distribution of W_{a^α, b^β} where $\alpha, \beta \in \{-, +\}$. Even though $W_{a,b}$ is in D -value representation, after a polarization transform this is no longer the case. Of course, one can always bring the polarized joint channel to an equivalent D -value representation as in Lemma 8.

The polar construction is shown in Figure 2, where we explicitly stated the different outputs of the polarized channels. We note that the top copy of $W_{a,b}$ outputs, jointly, $(y_a, u_a \oplus \nu_a, d_b)$, as its a -input is $u_a \oplus \nu_a$.

The input u_{a^α} and output y_{a^α} of W_{a^α} are given by

$$u_{a^\alpha} = \begin{cases} u_a & \alpha = - \\ \nu_a & \alpha = +, \end{cases}$$

$$y_{a^\alpha} = \begin{cases} (y_a, \eta_a) & \alpha = - \\ (y_a, \eta_a, u_a) & \alpha = +. \end{cases}$$

The input u_{b^β} and output y_{b^β} of W_{b^β} are given by

$$u_{b^\beta} = \begin{cases} u_b & \beta = - \\ \nu_b & \beta = +, \end{cases}$$

$$y_{b^\beta} = \begin{cases} (y_a, \eta_a, u_a, \nu_a, d_b, \delta_b) & \beta = - \\ (y_a, \eta_a, u_a, \nu_a, u_b, d_b, \delta_b) & \beta = +. \end{cases}$$

Note that y_{a^α} and u_{a^α} are contained in y_{b^β} . Thus, the joint output of both channels is y_{b^β} .

The distribution of the jointly polarized channel is given by

$$\begin{aligned} W_{a^\alpha, b^\beta}(y_{a^\alpha}, y_{b^\beta} | u_{a^\alpha}, u_{b^\beta}) \\ &= 2W_{b^\beta}(y_{b^\beta} | u_{b^\beta}) \\ &= \sum_{B_\beta} \left(W_b(y_a, u_a \oplus \nu_a, d_b | u_b \oplus \nu_b) W_b(\eta_a, \nu_a, \delta_b | \nu_b) \right), \end{aligned} \quad (17)$$

where

$$\sum_{B_\beta} \equiv \begin{cases} \sum_{\nu_b} & \beta = - \\ \text{No sum} & \beta = +. \end{cases}$$

We have shown how to generate W_{a^α, b^β} from $W_{a,b}$. Another case of interest is generating W_{a^-, a^+} from W_a . Denote the output of W_{a^-} by y_{a^-} . The output of W_{a^+} is (y_{a^-}, u_a) . From (8), we need only compute W_{a^+} to find W_{a^-, a^+} . This is accomplished by (1).

If two channels are ordered by degradation, so are their polar transforms [3, Lemma 4.7]. I.e., if $Q \succcurlyeq W$ then $Q^- \succcurlyeq W^-$ and $Q^+ \succcurlyeq W^+$. This is readily extended to joint channels.

Lemma 9. Let BMS channel $Q \succcurlyeq W$. Then $Q_{-,+} \stackrel{p}{\succcurlyeq} W_{-,+}$.

Proof: Using (4) and the definition of $W_{-,+}$ we have

$$\begin{aligned} W_{-,+}((y_1, y_2), u_1 | u_1, u_2) \\ &= \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2) \\ &= \sum_{z_1, z_2} \frac{1}{2} Q(z_1 | u_1 \oplus u_2) P(y_1 | z_1) Q(z_2 | u_2) P(y_2 | z_2) \\ &= \sum_{z_1, z_2} Q_{-,+}((z_1, z_2), u_1 | u_1, u_2) P_a(y_1, y_2 | z_1, z_2), \end{aligned}$$

where $P_a(y_1, y_2 | z_1, z_2) = P(y_1 | z_1) P(y_2 | z_2)$ is a proper degrading channel. ■

Lemma 10. If $Q_{a,b}(z_a, z_b | u_a, u_b) \stackrel{p}{\succcurlyeq} W_{a,b}(y_a, y_b | u_a, u_b)$, then, for $\alpha, \beta \in \{-, +\}$, $Q_{a^\alpha, b^\beta} \stackrel{p}{\succcurlyeq} W_{a^\alpha, b^\beta}$.

Proof: The proof follows similar lines to the proof of Lemma 9. Expand W_{a^α, b^β} using (17) and expand again using the definition of joint degradation with a proper degrading channel. Using the one-to-one mappings between the outputs of the polarized channels and the inputs and outputs of non-polarized channels, the desired results are obtained. The details are mostly technical, and are omitted. ■

The operational meaning of Lemma 10 is that to compute an upgraded approximation of W_{a^α, b^β} we may start with $Q_{a,b}$, an upgraded approximation of $W_{a,b}$, and polarize it. The result Q_{a^α, b^β} is an upgraded approximation of W_{a^α, b^β} . This enables us to iteratively compute upgraded approximations of joint synthetic channels. Whenever the joint synthetic channel exceeds an allotted size, we upgrade it to a joint channel with a smaller alphabet size and continue from there. We make sure to use proper upgrading procedures; this preserves the special structure of the joint channel and enables us to compute a lower bound on the probability of error. In Section VI we derive such upgrading procedures.

Since a sequence of polarization and upgrading steps is equivalent to upgrading the overall polarized joint distribution, using Lemmas 6 and 7 we obtain that the IMJP decoding error of a joint distribution that has undergone multiple polarization and proper upgrading steps lower-bounds the SC decoding error of the joint distribution that has undergone only the same polarization steps (without upgrading steps).

C. Double Symmetry for Joint Distributions

A binary input channel $W(y|u)$ is called *symmetric* if for every output y there exists a conjugate output \bar{y} such that $W(y|0) = W(\bar{y}|1)$. We now extend this to joint synthetic channels.

Definition 3 (Double symmetry). Joint channel $W_b(y_a, u_a, d_b|u_b)$ exhibits double symmetry if for every y_a, d_b there exist $y_a^{(a)}, y_a^{(b)}, y_a^{(ab)}$ such that

$$\begin{aligned} W_b(y_a, u_a, d_b|u_b) &= W_b(y_a^{(a)}, \bar{u}_a, d_b|u_b) \\ &= W_b(y_a^{(b)}, u_a, -d_b|\bar{u}_b) \\ &= W_b(y_a^{(ab)}, \bar{u}_a, -d_b|\bar{u}_b). \end{aligned} \quad (18)$$

We call $(\cdot)^{(a)}$ the a-conjugate; $(\cdot)^{(b)}$ the b-conjugate; and $(\cdot)^{(ab)}$ the ab-conjugate. We can also cast this definition using the regular (non- D -value) representation of joint channels in a straight-forward manner, which we omit here.

Example 3. Let W be a BMS channel, and consider the joint channel formed by its ‘-’- and ‘+’-transforms, $W_{-,+}$. What are the a-, b-, and ab-conjugates of the a-channel output y_a ? Recall that the output of the a-channel W^- consists of the outputs of two copies of W . Denote $y_a = (y_1, y_2)$, where y_1 and y_2 are two possible outputs of W with conjugates \bar{y}_1, \bar{y}_2 , respectively. We then have

$$\begin{aligned} W_{-,+}(y_a, u_a|u_b) &= 2W^+(y_a, u_a|u_b) \\ &= W(y_1|u_a \oplus u_b)W(y_2|u_b). \end{aligned}$$

By symmetry of W we obtain $y_a^{(a)} = (\bar{y}_1, y_2)$, $y_a^{(b)} = (\bar{y}_1, \bar{y}_2)$, and $y_a^{(ab)} = (y_1, \bar{y}_2)$. Indeed,

$$\begin{aligned} W^+(y_a, u_a|u_b) &= W^+(y_a^{(a)}, \bar{u}_a|u_b) \\ &= W^+(y_a^{(b)}, u_a|\bar{u}_b) \\ &= W^+(y_a^{(ab)}, \bar{u}_a|\bar{u}_b). \end{aligned}$$

We leave it to the reader to show that (18) holds for the D -value representation of the joint channel.

Pairs of polar synthetic channels exhibit double symmetry. One can see this directly from symmetry properties of polar synthetic channels, see [1, Proposition 13]. Alternatively, one can use induction to show directly that the polar construction preserves double symmetry; we omit the details. This implies the following Proposition.

Proposition 11. Let $W_{a,b}$ be the joint distribution of two synthetic channels W_a and W_b that result from n polarization steps of BMS channel W . Then, $W_{a,b}$ exhibits double symmetry.

The following is a direct consequence of double symmetry.

Lemma 12. Let $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ be a joint distribution in D -value representation that exhibits double symmetry. Then

- 1) For the b-channel, (y_a, u_a, d_b) and $(y_a^{(a)}, \bar{u}_a, d_b)$ have the same b-channel D -value d_b .
- 2) For the a-channel, y_a and $y_a^{(b)}$ have the same a-channel D -value d_a , and $y_a^{(a)}$ and $y_a^{(ab)}$ have the same a-channel D -value $-d_a$.

Proof: The first item is obvious from (18). For the second item, note that

$$\begin{aligned} W_a(y_a|u_a) &= \sum_{d_b} \sum_{u_b} W_b(y_a, u_a, d_b|u_b) \\ &\stackrel{(a)}{=} \sum_{d_b} \sum_{u_b} W_b(y_a^{(b)}, u_a, -d_b|\bar{u}_b) \\ &= \sum_{-d_b} \sum_{\bar{u}_b} W_b(y_a^{(b)}, u_a, -d_b|\bar{u}_b) \\ &= W_a(y_a^{(b)}|u_a), \end{aligned}$$

where (a) is by (18). In the same manner, $y_a^{(a)}$ and $y_a^{(ab)}$ have the same a-channel D -value, $-d_a$. ■

Lemma 12 implies that an SC decoder does not distinguish between y_a and $y_a^{(b)}$ when making its decision for the a-channel. We now show that a similar conclusion holds for the IMJP decoder.

Lemma 13. Let y_a be some output of W_a . Then

$$T(y_a|u_a) = T(y_a^{(b)}|u_a) = T(y_a^{(a)}|\bar{u}_a) = T(y_a^{(ab)}|\bar{u}_a).$$

Proof: Theorem 1 holds for joint channels given in D -value representation, $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$. This is easily seen by following the proof with minor changes. Under the D -value representation, (11) becomes

$$\begin{aligned} T(y_a|u_a) &= \frac{1}{2} \sum_{d_b} \max_{u_b} W_{a,b}(y_a, u_a, d_b|u_a, u_b) \\ &= \sum_{d_b} \max_{u_b} W_b(y_a, u_a, d_b|u_b). \end{aligned} \quad (19)$$

The remainder of the proof hinges on double symmetry and follows along similar lines to the proof of Lemma 12, with W_a replaced with T and accordingly the sum over u_b replaced with a maximum operation over u_b . ■

Lemma 13 implies that the IMJP decoder does not distinguish between y_a and $y_a^{(b)}$.

Corollary 14. Let ϕ_a be the IMJP decoder for the a-channel. Then $\phi_a(y_a) = \phi_a(y_a^{(b)}) = 1 - \phi_a(y_a^{(a)}) = 1 - \phi_a(y_a^{(ab)})$.

V. SYMMETRIZED JOINT SYNTHETIC CHANNELS

In this section we introduce the symmetrizing transform. The resultant channel is *degraded* from the original joint distribution yet has the same probability of error. Its main merit is to decouple the a-channel from the b-channel. This simpler structure is the key to upgrading the a-channel, as we shall see in Section VI.

A. Symmetrized Joint Distribution

The SC decoder observes marginal distributions and makes a decision based on the D -value of each synthetic channel's output. In particular, by Lemma 12, the SC decoder makes the same decision for the a-channel whether its output was y_a or $y_a^{(b)}$ and the b-channel decision is based on d_b without regard to y_a . By Corollary 14, the IMJP decoder acts similarly. That is, the IMJP decoder makes the same decision for the

a-channel whether its output is y_a or $y_a^{(b)}$, and the decision for the b-channel is based solely on d_b .

We conclude that if the a-channel were told only whether its output was one of $\{y_a, y_a^{(b)}\}$, it would make the same decision had it been told its output was, say, y_a . This is true for either the SC or IMJP decoder. Consequently, either decoder's probability of error is unaffected by obscuring the a-channel output in this manner.

This leads us to define a *symmetrized* version of the joint synthetic channel distribution, $\hat{W}_{a,b}$, as follows. Let⁴

$$\begin{aligned}\hat{y}_a &\triangleq \{y_a, y_a^{(b)}\}, \\ \bar{\hat{y}}_a &\triangleq \{y_a^{(a)}, y_a^{(ab)}\}\end{aligned}$$

and define

$$\begin{aligned}\hat{W}_{a,b}(\hat{y}_a, u_a, d_b | u_a, u_b) &= W_{a,b}(y_a, u_a, d_b | u_a, u_b) \\ &\quad + W_{a,b}(y_a^{(b)}, u_a, d_b | u_a, u_b), \\ \hat{W}_{a,b}(\bar{\hat{y}}_a, u_a, d_b | u_a, u_b) &= W_{a,b}(y_a^{(a)}, u_a, d_b | u_a, u_b) \\ &\quad + W_{a,b}(y_a^{(ab)}, u_a, d_b | u_a, u_b).\end{aligned}\tag{20}$$

Lemma 15. *Let $W_{a,b}$ be a joint synthetic channel distribution, and let $\hat{W}_{a,b}$ be its symmetrized version. Then, the probability of error under SC (IMJP) decoding of either channel is identical.*

Proof: By Lemma 12 for the SC decoder or Corollary 14 for the IMJP decoder, if the decoder for the symmetrized channel makes an error for some symbol \hat{y}_a then the decoder for the non-symmetrized channel make an error for both y_a and $y_a^{(b)}$, and vice-versa. Therefore, denoting by \mathcal{E} the error indicator of the decoder,

$$\begin{aligned}P_e(\hat{W}_{a,b}) &= \frac{1}{4} \sum_{u_a, u_b} \sum_{\hat{y}_a, d_b} \hat{W}_{a,b}(\hat{y}_a, u_a, d_b | u_a, u_b) \mathcal{E} \\ &\stackrel{(a)}{=} \frac{1}{4} \sum_{u_a, u_b} \sum_{y_a, d_b} W_{a,b}(y_a, u_a, d_b | u_a, u_b) \mathcal{E} \\ &= P_e(W_{a,b}),\end{aligned}$$

where (a) is by (20). \blacksquare

The marginal synthetic channels \hat{W}_a and \hat{W}_b are given by

$$\begin{aligned}\hat{W}_a(\hat{y}_a | u_a) &= \sum_{u_b, d_b} \hat{W}_{a,b}(\hat{y}_a, u_a, d_b | u_a, u_b), \\ \hat{W}_b(\hat{y}_a, u_a, d_b | u_b) &= \frac{1}{2} \hat{W}_{a,b}(\hat{y}_a, u_a, d_b | u_a, u_b).\end{aligned}$$

Note that by double symmetry

$$\begin{aligned}\hat{W}_a(\hat{y}_a | u_a) &= \hat{W}_a(\bar{\hat{y}}_a | \bar{u}_a), \\ \hat{W}_b(\hat{y}_a, u_a, d_b | u_b) &= \hat{W}_b(\bar{\hat{y}}_a, \bar{u}_a, d_b | u_b) \\ &= \hat{W}_b(\hat{y}_a, u_a, -d_b | \bar{u}_b) \\ &= \hat{W}_b(\bar{\hat{y}}_a, \bar{u}_a, -d_b | \bar{u}_b).\end{aligned}\tag{21}$$

Definition 4 (Symmetrized distribution). A joint distribution whose marginals satisfy (21) is called *symmetrized*.

⁴The order of elements in \hat{y}_a and $\bar{\hat{y}}_a$ does not matter. I.e., $\{y_a, y_a^{(b)}\}$ is a set containing both y_a and $y_a^{(b)}$.

The name ‘symmetrized’ stems from comparison of (21) and (18). We note that Theorem 1 holds for $\hat{W}_{a,b}$.

A symmetrized joint distribution remains symmetrized upon polarization. That is, if $\hat{W}_{a,b}$ is a symmetrized joint distribution and $\hat{W}_{a^\alpha, b^\beta}$, $\alpha, \beta \in \{-, +\}$ is the result of jointly polarizing it, then the marginals \hat{W}_{a^α} and \hat{W}_{b^β} satisfy (21). This is easily seen from (17) and (21).

Clearly, $\hat{W}_{a,b}$ is *degraded* with respect to $W_{a,b}$, exactly the opposite of our main thrust. Nevertheless, as established in Lemma 15, both channels have the same probability of error under SC (IMJP) decoding. Moreover, if we upgrade the symmetrized version of the channel, its probability of error under IMJP decoding lower-bounds the probability of error of the non-symmetrized channel under either SC or IMJP decoding.

What isn’t immediately obvious, however, is what happens after polarization. I.e., if we take a joint channel, symmetrize it, and then polarize it, how does its probability of error compare to the original joint channel that has just undergone polarization? Furthermore, what happens if the symmetrized version undergoes an upgrading transform?

Proposition 16. *Let $W_{a,b}$ be a joint distribution of two synthetic channels and let $W_{a,b}^t$ denote this joint distribution after a sequence t of joint polarization steps. Then $P_e^{\text{IMJP}}(W_{a,b}^t) \geq P_e^{\text{IMJP}}(\hat{Q}_{a,b}^t)$, where $\hat{Q}_{a,b}^t$ is the distribution of $\hat{W}_{a,b}$ after the same sequence of polarization steps and any number of proper upgrading transforms along the way.*

Proof: Let W_{a^α, b^β} and $\hat{W}_{a^\alpha, b^\beta}$ be the polarized versions of $W_{a,b}$ and $\hat{W}_{a,b}$, respectively. For the b^β -channel, the decoder makes the same decision for either W_{a^α, b^β} or $\hat{W}_{a^\alpha, b^\beta}$. This is because the decision is based on the b-channel D -value, which is unaffected by symmetrization (see (20)).

Next, for the a^α channel, using on (17) a derivation similar to the proof of Lemma 13, $T(y_{a^\alpha} | u_{a^\alpha}) = T(y'_{a^\alpha} | u_{a^\alpha})$, where y'_{a^α} is any combination of an element of \hat{y}_a and an element of $\hat{\eta}_a$. I.e., y'_{a^α} is any one of $\{y_a, \eta_a\}$, $\{y_a^{(b)}, \eta_a\}$, $\{y_a, \eta_a^{(b)}\}$, $\{y_a^{(b)}, \eta_a^{(b)}\}$. Thus, the IMJP decoder makes the same decision for the a^α -channel for either W_{a^α, b^β} or $\hat{W}_{a^\alpha, b^\beta}$.

We compare the channels obtained by the following two procedures.

- *Procedure 1:* Joint channel $W_{a,b}$ goes through sequence t of polarization steps.
- *Procedure 2:* Joint channel $W_{a,b}$ is symmetrized to form $\hat{W}_{a,b}$. It goes through sequence t of polarization steps (without any further symmetrization operations).

We iteratively apply the above reasoning and conclude in a similar manner to Lemma 15 that both channels have the same performance under IMJP decoding. Next, we modify Procedure 2.

- *Procedure 2a:* Joint channel $W_{a,b}$ is symmetrized to form $\hat{W}_{a,b}$. It goes through sequence t of polarization steps (without any further symmetrization operations), but at some point mid-sequence, it undergoes a proper upgrading procedure.

Since polarizing and proper upgrading is equivalent to proper upgrading and polarizing, Lemma 10, we can assume that the upgrading happens after the entire sequence of polarization steps. Thus, under IMJP decoding, the probability of error of the channel that results from Procedure 2a lower-bounds the probability of error of the channels resulting from Procedures 1 and 2. Similarly, multiple upgrading transforms can also be thought of as occurring after all polarization steps. ■

Corollary 17. *Let W be a BMS channel that undergoes n polarization steps. Let $W_{a,b}$ be the joint distribution of two of its polar descendants, and let $\hat{Q}_{a,b} \stackrel{p}{\succ} \hat{W}_{a,b}$. Then $P_e^{\text{SC}}(W) \geq P_e^{\text{IMJP}}(\hat{Q}_{a,b})$.*

Proof: A direct consequence of Lemmas 6 and 7 combined with Proposition 16. ■

Due to Proposition 16, we henceforth assume that joint channel $W_{a,b}$ is symmetrized, and no longer distinguish symmetrized channels or symbols by the (\circ) symbol. Replacing the joint channel with its symmetrized version need only be performed once, at the first instance the two channels go through different polarization transforms.

Implementation: Since symmetrization is performed only once, and since this invariably happens when converting a channel W to $W_{-,+}$, we find the a -, b -, and ab -conjugates using the results of Example 3. We then form the symmetrized channel using (20). Note that it is sufficient to find just the b -conjugates and use the first equation of (20).

B. Decomposition of Symmetrized Joint Distributions

Let the joint channel be $W_b(y_a, u_a, d_b|u_b)$, which, as mentioned above, we assume to be symmetrized. We have

$$\begin{aligned} W_b(y_a, u_a, d_b|u_b) &= \mathbb{P}\{y_a, u_a|u_b\} \mathbb{P}\{d_b|u_b, y_a, u_a\} \\ &= \mathbb{P}\{u_a\} \mathbb{P}\{y_a|u_a, u_b\} \mathbb{P}\{d_b|u_b, y_a, u_a\} \\ &= \frac{1}{2} W_1(y_a|u_a, u_b) \cdot W_2(d_b|u_b; y_a, u_a), \end{aligned} \quad (22)$$

in which we used the independence and uniformity of the input bits u_a and u_b . The distribution W_1 is given by $W_1(y_a|u_a, u_b) = 2 \sum_{d_b} W_b(y_a, u_a, d_b|u_b)$. Whenever $W_1(y_a|u_a, u_b)$ is nonzero, distribution $W_2(d_b|u_b; y_a, u_a)$ is obtained by dividing $W_b(y_a, u_a, d_b|u_b)$ by $W_1(y_a|u_a, u_b)/2$. Our notation $W_2(d_b|u_b; y_a, u_a)$ (with a semicolon, as opposed to $W_2(d_b|y_a, u_a, d_b)$) reminds us that for fixed y_a, u_a , W_2 is binary-input channel with input u_b and output d_b . If $W_1(y_{a0}|u_{a0}, u_b) = 0$ for some y_{a0}, u_{a0} , we define $W_2(d_b|u_b; y_{a0}, u_{a0})$ to be some arbitrary BMS channel, to ensure it is always a valid channel.

Since the joint channel is symmetrized, by (21) we have $W_1(y_a|u_a, u_b) = W_1(y_a|u_a, \bar{u}_b)$. Hence, for any u_b ,

$$W_a(y_a|u_a) = \sum_{u'_b} W_1(y_a|u_a, u'_b) \mathbb{P}\{u'_b\} = W_1(y_a|u_a, u_b). \quad (23)$$

I.e., a consequence of symmetrization is that given u_a, y_a becomes *independent* of u_b . This is not true in the general case where the joint channel is not symmetrized.

The decomposition of (22) essentially decouples the symmetrized joint channel to a product of two distributions.

Lemma 18. *Let $W_b(y_a, u_a, d_b|u_b)$ be a symmetrized joint distribution. It admits the decomposition*

$$W_b(y_a, u_a, d_b|u_b) = \frac{1}{2} W_a(y_a|u_a) W_2(d_b|u_b; y_a, u_a). \quad (24)$$

For any y_a, u_a , W_2 is a BMS channel with input u_b and output d_b , i.e.,

$$W_2(d_b|u_b; y_a, u_a) = W_2(-d_b|\bar{u}_b; y_a, u_a).$$

Moreover, W_2 satisfies

$$W_2(d_b|u_b; y_a, u_a) = W_2(d_b|u_b; \bar{y}_a, \bar{u}_a). \quad (25)$$

Proof: Using (23) in (22) yields (24). The remainder of this lemma is readily obtained by using (21) in (24). ■

Definition 5 (Decoupling decomposition). A decomposition of the form (24) for a symmetrized joint distribution is called a *decoupling decomposition*. Channel W_a is obtained by marginalization, $W_a(y_a|u_a) = 2 \sum_{d_b} W_b(y_a, u_a, d_b|u_b)$ for any u_b . Channel $W_2(d_b|u_b; y_a, u_a)$ is obtained by dividing $W_b(y_a, u_a, d_b|u_b)$ by $W_a(y_a|u_a)/2$ if $W_a(y_a|u_a)$ is nonzero, and set to an arbitrary BMS channel, e.g., some BSC, if $W_a(y_a|u_a) = 0$.⁵ When setting to an arbitrary channel, we make sure not to add new b -channel D -values.

We use decoupling decompositions of symmetrized joint distributions in the sequel.

VI. UPGRADING PROCEDURES FOR JOINT SYNTHETIC CHANNELS

In this section, we introduce proper upgrading procedures for joint synthetic channels. The overall goal is to reduce the alphabet size of the joint distribution. The upgrading procedures we develop enable us to reduce the alphabet size of each of the marginals without changing the distribution of the other; there is a different procedure for each marginal. As an intermediate step, we further couple the marginals by increasing the alphabet size of one of them.

The joint channel $W_{a,b}$ is assumed to be symmetrized and in D -value representation. The upgrading procedures will maintain this. As discussed in Section V, we do not distinguish symmetrized channels with any special symbol. The upgrading procedure of Section VI-A hinges on symmetrization. The upgrading procedure of Section VI-B does not require symmetrization and holds for non-symmetrized channels without change. However, we shall see that symmetrization simplifies the resulting expressions.

A. Upgrading Channel W_a

We now introduce a theorem that enables us to deduce an upgrading procedure that upgrades W_a and reduces its output alphabet size. Let symmetrized joint channel $W_b(y_a, u_a, d_b|u_b)$ admit decoupling decomposition (24). Let $Q_b(z_a, u_a, z_b|u_b)$ be

⁵This invariably happens for perfect symbols. I.e., symbols for which $W_a(y_a|u_a) > 0$ but $W_a(y_a|\bar{u}_a) = 0$ for some $u_a \in \{0, 1\}$.

another symmetrized joint channel, where z_b represents the D -value of the b-channel output. It also admits a decoupling decomposition,

$$Q_b(z_a, u_a, z_b|u_b) = \frac{1}{2} Q_a(z_a|u_a) Q_2(z_b|u_b; z_a, u_a). \quad (26)$$

Theorem 19. Let W_b and Q_b be symmetrized joint distributions with decoupling decompositions (24) and (26), respectively.

Then, $Q_b \stackrel{p}{\succ} W_b$ if

- 1) Channel $Q_a(z_a|u_a) \succcurlyeq W_a(y_a|u_a)$ with degrading channel $P_a(y_a|z_a)$.
- 2) Channel $Q_2(z_b|u_b; z_a, u_a) \succcurlyeq W_2(d_b|u_b; y_a, u_a)$ for all u_a, y_a, z_a such that $P_a(y_a|z_a) > 0$.

Before going into the proof, some comments are in order. First, we do not claim that any Q_b that is upgraded from W_b must satisfy this theorem. Second, the meaning of the second item is that, for fixed z_a, u_a , BMS channel $Q_2(z_b|u_b; z_a, u_a)$ with binary input u_b is upgraded from a set of BMS channels $\{W_2(d_b|u_b; y_a, u_a)\}_{y_a}$ with the same binary input.

Proof: Using decoupling decompositions (24) and (26) and the structure of a proper degrading channel (16), channel $Q_b \stackrel{p}{\succ} W_b$ if and only if there exist P'_a, P'_b such that

$$\sum_{z_a} Q_a(z_a|u_a) P'_a(y_a|z_a) V(d_b|z_a, y_a, u_a, u_b) = W_a(y_a|u_a) W_2(d_b|u_b; y_a, u_a), \quad (27)$$

where

$$V(d_b|z_a, y_a, u_a, u_b) = \sum_{z_b} Q_2(z_b|u_b; z_a, u_a) P'_b(d_b|y_a, z_a, u_a, z_b). \quad (28)$$

We now find P'_a and P'_b from the conditions of the theorem.

The first condition of the theorem implies that there exists a channel $P_a(y_a|z_a)$ such that

$$\sum_{z_a} Q_a(z_a|u_a) P_a(y_a|z_a) = W_a(y_a|u_a).$$

The second condition of the theorem implies that for each y_a, u_a, z_a there exists a channel $P_b(d_b|y_a, z_a, u_a, z_b)$ such that

$$\sum_{z_b} Q_2(z_b|u_b; z_a, u_a) P_b(d_b|y_a, z_a, u_a, z_b) = W_2(d_b|u_b; y_a, u_a) \cdot \mathbb{P}\{y_a|z_a\} > 0. \quad (29)$$

We set

$$P'_a(y_a|z_a) = P_a(y_a|z_a), \\ P'_b(d_b|y_a, z_a, u_a, z_b) = P_b(d_b|y_a, z_a, u_a, z_b).$$

Using (29) in (28), we have

$$V(d_b|z_a, y_a, u_a, u_b) = W_2(d_b|u_b; y_a, u_a) \cdot \mathbb{P}\{P_a(y_a|z_a) > 0\}.$$

It is easily verified that (27) is satisfied by $P'_a = P_a$ and this V , completing the proof. ■

How might one use Theorem 19 to upgrade the a-channel? A naive way would be to first upgrade the marginal W_a to Q_a using some known method (e.g., the methods of [11], see

Appendix C). This yields degrading channel P_a by which one can find channel Q_2 that satisfies (29). With Q_a and Q_2 at hand, one forms the product (26) to obtain Q_b . If the reader were to attempt to do this, she would find out that it often changes the b-channel. Moreover, this change may be radical: the resulting b-channel may be so upgraded to become almost noiseless, which boils down to an uninteresting bound, the trivial lower bound (2). It is possible to upgrade the a-channel without changing the b-channel; this requires an additional transform we now introduce.

The *upgrade-couple* transform enables upgrading the a-channel without changing the b-channel. The idea is to split each a-channel symbol to several classes, according to the possible b-channel outputs. Symbols within a class have the same W_2 channel, so that confining upgrade-merges to operate within a class inherently satisfies the second condition of Theorem 19. Thus, we circumvent changes to the b-channel. This results in only a modest increase to the number of output symbols of the overall joint distribution.

Let channel W_b have $2B$ possible D -values, $\pm d_{b1}, \pm d_{b2}, \dots, \pm d_{bB}$. We assume that erasure symbols are duplicated,⁶ and $0 \leq d_{b1} \leq d_{b2} \leq \dots \leq d_{bB} \leq 1$. For each a-channel symbol y_a we define B^2 upgrade-couple symbols $y_a^{i,j}$, $i, j \in \{1, 2, \dots, B\}$. The new symbols *couple* the outputs of the a- and b-channels (whence the name of the upgrade-couple transform). Namely, if the a-channel output is $y_a^{i,j}$ and $u_a = 0$, the b-channel output can only be $\pm d_{bi}$; if the a-channel output is $y_a^{i,j}$ and $u_a = 1$, the b-channel output can only be $\pm d_{bj}$.

The upgrade-couple channel $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ is defined by

$$\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) \triangleq W_b(y_a, u_a, d_b|u_b) \cdot S_{i,j}(y_a, u_a, d_b), \quad (30)$$

where

$$S_{i,j}(y_a, u_a, d_b) = \begin{cases} \sum_{d=\pm d_{bj}} W_2(d|u_b; y_a, \bar{u}_a) & u_a = 0, \\ & d_b = \pm d_{bi} \\ \sum_{d=\pm d_{bi}} W_2(d|u_b; y_a, \bar{u}_a) & u_a = 1, \\ & d_b = \pm d_{bj} \\ 0 & \text{otherwise,} \end{cases}$$

and W_2 is from the decoupling decomposition of W_b in (24). The factor $S_{i,j}(y_a, u_a, d_b)$ is indeed independent of u_b due to symmetry.

As we now show, since W_b is symmetrized, so is \check{W}_b .

Lemma 20. Let $W_b(y_a, u_a, d_b|u_b)$ be a symmetrized joint distribution. Then, $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$, defined as in (30), is also symmetrized.

Proof: To establish the lemma, we need to show that (20) holds for the upgrade-couple channel. For the a-channel W_a , let symbols y_a, \bar{y}_a be conjugates, i.e., $W_a(y_a|u_a) = W_a(\bar{y}_a|\bar{u}_a)$. Channel W_b is symmetrized, so it satisfies (20) under which

⁶I.e., there is a “positive” and a “negative” erasure, see [11, Lemma 4].

$S_{i,j}(y_a, u_a, d_b) = S_{j,i}(\bar{y}_a, \bar{u}_a, d_b)$. Furthermore, obviously $S_{i,j}(y_a, u_a, d_b) = S_{i,j}(y_a, u_a, -d_b)$. Thus,

$$\begin{aligned}\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) &= \check{W}_b(\bar{y}_a^{j,i}, \bar{u}_a, d_b|u_b) \\ &= \check{W}_b(y_a^{i,j}, u_a, -d_b|\bar{u}_b) \\ &= \check{W}_b(\bar{y}_a^{j,i}, \bar{u}_a - d_b|\bar{u}_b).\end{aligned}$$

Next, recall that $\check{W}_a(y_a^{i,j}|u_a) = \sum_{d_b, u_b} \check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$, so that $\check{W}_a(y_a^{i,j}|u_a) = \check{W}_a(\bar{y}_a^{j,i}|\bar{u}_a)$. Thus, (20) holds as required. ■

In the proof of Lemma 20 we have seen that the conjugate symbol of $y_a^{i,j}$ is $\bar{y}_a^{j,i}$ (with the order of i and j flipped). We summarize this in the following corollary.

Corollary 21. *If $W_a(\bar{y}_a|\bar{u}_a) = W_a(y_a|u_a)$ then $\check{W}_a(\bar{y}_a^{j,i}|\bar{u}_a) = \check{W}_a(y_a^{i,j}|u_a)$.*

Since \check{W}_b is symmetrized, it admits decoupling decomposition

$$\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) = \frac{1}{2} \check{W}_a(y_a^{i,j}|u_a) \check{W}_2(d_b|u_b; y_a^{i,j}, u_a). \quad (31)$$

In Lemma 22 we derive \check{W}_a (see (33)) and establish that for every y_a ,

$$\check{W}_2(d_b|u_b; y_a^{i,j}, u_a) = \begin{cases} \text{BSC}\left(\frac{1-d_{bi}}{2}\right) & u_a = 0 \\ \text{BSC}\left(\frac{1-d_{bj}}{2}\right) & u_a = 1. \end{cases} \quad (32)$$

I.e., when $u_a = 0$ we have $\check{W}_2(\pm d_{bi}|u_b; y_a^{i,j}, u_a) = (1 \pm (-1)^{u_b} d_{bi})/2$, when $u_a = 1$ we have $\check{W}_2(\pm d_{bj}|u_b; y_a^{i,j}, u_a) = (1 \pm (-1)^{u_b} d_{bj})/2$, and $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ is zero for any other d_b . We remark that we define $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ using (32) even if $\check{W}_a(y_a^{i,j}|u_a) = 0$.

Lemma 22. *Let $W_b(y_a, u_a, d_b|u_b)$ be a symmetrized joint distribution and let $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ be defined as in (30), with decoupling decomposition (31). Then*

- 1) *Joint channel \check{W}_b is upgraded from joint channel W_b with a proper degrading channel that deterministically maps $y_a^{i,j}$ to y_a .*
- 2) *Symbols y_a of channel W_a and $y_a^{i,j}$ of channel \check{W}_a have the same a-channel D -value for every i, j such that $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) > 0$.*
- 3) *For every y_a , BMS channel $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ with input u_b and output d_b is BSC($(1-d_{bi})/2$) if $u_a = 0$ and BSC($(1-d_{bj})/2$) if $u_a = 1$.*

Proof: For the first item, note that $\sum_{i,j} S_{i,j}(y_a, u_a, d_b) = 1$. By summing (30) over i, j we obtain $\check{W}_b(y_a, u_a, d_b|u_b) = \sum_{i,j} \check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$. I.e., joint channel \check{W}_b is upgraded from W_b with degrading channel P_a that deterministically maps $y_a^{i,j}$ to y_a . This is a proper degrading channel.

For the second item, we marginalize \check{W}_b over d_b and u_b and obtain that for every u_a ,

$$\check{W}_a(y_a^{i,j}|u_a) = W_a(y_a|u_a) \cdot \alpha_{i,j}(y_a), \quad (33)$$

where

$$\alpha_{i,j}(y_a) = S_{i,j}(y_a, 0, d_{bi}) \cdot S_{i,j}(y_a, 1, d_{bj}).$$

Whenever $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) > 0$, we have $\alpha_{i,j}(y_a) > 0$. Thus,

$$\frac{\check{W}_a(y_a^{i,j}|0) - \check{W}_a(y_a^{i,j}|1)}{\check{W}_a(y_a^{i,j}|0) + \check{W}_a(y_a^{i,j}|1)} = \frac{W_a(y_a|0) - W_a(y_a|1)}{W_a(y_a|0) + W_a(y_a|1)},$$

implying that y_a and $y_a^{i,j}$ have the same a-channel D -value for their respective channels.

For the final item, if $\check{W}_a(y_a^{i,j}|u_a) = 0$, we are free to set $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ as we please, so we set it as per the item. Otherwise, there are only two values of d_b for which $S_{i,j}(y_a, u_a, d_b)$ is nonzero. Hence, \check{W}_b can output only two b-channel D -values for fixed $y_a^{i,j}$ and u_a . Thus, \check{W}_2 is a BMS channel with only two possible outputs, or, in other words, a BSC. A BSC that outputs D -values $\pm d$, $0 \leq d \leq 1$, has crossover probability $(1-d)/2$. This establishes the item. ■

Definition 6 (Canonical channel). The canonical channel $W^*(d|u)$ of channel $W(y|u)$ has a single entry for each D -value. I.e., denoting by D_d the set of symbols y whose D -value is d , we have $W^*(d|u) = \sum_{D_d} W(y|u)$. It can be shown that a channel is equivalent to its canonical form, i.e., each form can be degraded from the other.

Corollary 23. *The canonical b-channels of $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ and $W_b(y_a, u_a, d_b|u_b)$ coincide.*

Proof: This is a direct consequence of the first item of Lemma 22:

$$\begin{aligned}\check{W}_b^*(d_b|u_b) &= \sum_{y_a, u_a} \sum_{i,j} \check{W}_b(y_a^{i,j}, u_a, d_b|u_b) \\ &= \sum_{y_a, u_a} W_b(y_a, u_a, d_b|u_b) \\ &= W_b^*(d_b|u_b).\end{aligned} \quad \blacksquare$$

Corollary 24. *The canonical a-channels of $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ and $W_b(y_a, u_a, d_b|u_b)$ coincide.*

Proof: A direct consequence of the second item of Lemma 22, using (33), and noting that $\sum_{i,j} \alpha_{i,j}(y_a) = 1$ for any y_a . ■

Definition 7 (Class). The class $C_{i,j}$ is the set of symbols $y_a^{i,j}$ with fixed i, j .

There are B^2 classes. The size of each class is the number of symbols y_a . By (32), $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ is the same BSC for all symbols of class $C_{i,j}$ and fixed u_a . Thus, the second item of Theorem 19 becomes trivial and is immediately satisfied if we use an upgrading procedure that upgrade-merges several symbols of the same class $C_{i,j}$.

To determine which upgrading procedures may be used, we turn to the degrading channel. So long as the degrading channel does not mix a symbol and its conjugate, the upgrading procedure can be confined to a single class. This is because conjugate symbols belong to different classes, as established in Corollary 21. Thus, of the upgrading procedures of [11] we can use either upgrade-merge-3 without restriction or upgrade-merge-2 provided that the two symbols to be merged have the same a-channel D -value.

Theorem 25. Let $W_b(y_a, u_a, d_b|u_b)$ be some joint distribution with marginals $W_a(y_a|u_a)$, $W_b^*(d_b|u_b)$ and upgrade-couple counterpart $\tilde{W}_b(y_a^{i,j}, u_a, d_b|u_b)$. Let $Q_a(z_a|u_a) \succcurlyeq W_a(y_a|u_a)$ obtained by an upgrade-merge-3 procedure. Then there exists joint distribution $\tilde{Q}_b(z_a^{i,j}, u_a, d_b|u_b) \succcurlyeq^p \tilde{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ with canonical marginals $\tilde{Q}_a^*(z_a|u_a), \tilde{Q}_b^*(d_b|u_b)$ such that $\tilde{Q}_a^* = Q_a^*$ and $\tilde{Q}_b^* = W_b^*$.

Proof: The idea is to confine the upgrading procedures to work within a class, utilizing Theorem 19 over each class separately.

Assume that the upgrading procedure from W_a to Q_a replaces symbols y_{a1}, y_{a2}, y_{a3} with symbols z_{a1}, z_{a3} . We obtain \tilde{Q}_b by using Theorem 19 for each class $C_{i,j}$ of \tilde{W}_b separately. The a-channel upgrade procedure for class $C_{i,j}$ is upgrade-merge-3 from \tilde{W}_a to \tilde{Q}_a that replaces symbols $y_{a1}^{i,j}, y_{a2}^{i,j}, y_{a3}^{i,j}$ with symbols $z_{a1}^{i,j}, z_{a3}^{i,j}$. As the upgrade is confined to symbols of the same class, the channel \tilde{W}_2 is the same regardless of y_a , as established in Lemma 22, item 3. Hence, within a class $C_{i,j}$ the second item of Theorem 19 is automatically satisfied with

$$\tilde{Q}_2(d_b|u_b; z_a^{i,j}, u_a) = \tilde{W}_2(d_b|u_b; y_a^{i,j}, u_a) \quad (34)$$

for all y_a, z_a . Channel \tilde{Q}_b is then obtained by the product of \tilde{Q}_a and \tilde{Q}_2 as per (31):

$$\tilde{Q}_b(z_a^{i,j}, u_a, d_b|u_b) = \frac{1}{2} \tilde{Q}_a(z_a^{i,j}|u_a) \tilde{Q}_2(d_b|u_b; z_a^{i,j}, u_a). \quad (35)$$

By properties of upgrade-merge-3 (see (49) in Appendix C-B) we have $\sum_{z_a} \tilde{Q}_a(z_a^{i,j}|u_a) = \sum_{y_a} \tilde{W}_a(y_a^{i,j}|u_a)$. Therefore,

$$\begin{aligned} \tilde{Q}_b^*(d_b|u_b) &= \sum_{i,j,u_a} \sum_{z_a} \tilde{Q}_b(z_a^{i,j}, u_a, d_b|u_b) \\ &\stackrel{(a)}{=} \sum_{i,j,u_a} \sum_{z_a} \frac{1}{2} \tilde{Q}_2(d_b|u_b; z_a^{i,j}, u_a) \tilde{Q}_a(z_a^{i,j}|u_a) \\ &\stackrel{(b)}{=} \sum_{i,j,u_a} \frac{1}{2} \tilde{W}_2(d_b|u_b; y_a^{i,j}, u_a) \sum_{z_a} \tilde{Q}_a(z_a^{i,j}|u_a) \\ &= \sum_{i,j,u_a} \frac{1}{2} \tilde{W}_2(d_b|u_b; y_a^{i,j}, u_a) \sum_{y_a} \tilde{W}_a(y_a^{i,j}|u_a) \\ &\stackrel{(c)}{=} \sum_{i,j,u_a} \sum_{y_a} \frac{1}{2} \tilde{W}_2(d_b|u_b; y_a^{i,j}, u_a) \tilde{W}_a(y_a^{i,j}|u_a) \\ &\stackrel{(d)}{=} W_b^*(d_b|u_b), \end{aligned}$$

where in (a) we used the decoupling decomposition (35); (b) and (c) are by Lemma 22, item 3 and by (34); finally, (d) is due to Corollary 23.

To see that the canonical a-channel marginals coincide, note that by Lemma 22, item 2, for any fixed z_a , the symbols $\{z_a^{i,j}\}_{i,j}$ all have the same a-channel D -value. Let d_a be some a-channel D -value, and let D_{d_a} be the set of a-channel outputs z_a whose a-channel D -value is d_a . Then,

$$\begin{aligned} \tilde{Q}_a^*(d_a|u_a) &= \sum_{z_a \in D_{d_a}} \sum_{i,j} \sum_{d_b, u_b} \tilde{Q}_b(z_a^{i,j}, u_a, d_b|u_b) \\ &= \sum_{z_a \in D_{d_a}} \sum_{i,j} \tilde{Q}_a(z_a^{i,j}|u_a) \end{aligned}$$

$$\begin{aligned} &\stackrel{(a)}{=} \sum_{z_a \in D_{d_a}} Q_a(z_a|u_a) \\ &= Q_a^*(d_a|u_a), \end{aligned}$$

where (a) is a direct consequence of the expressions for upgrade-merge-3 and our construction of upgrading each class separately. ■

To use Theorem 25, one begins with a design parameter A that controls the output alphabet size. Working one class at a time, one then applies upgrade operations in succession to reduce the class size to $2A$. The resulting channel, therefore, will have $2AB^2$ symbols overall. The canonical a-channel marginal that results from this operation will have at most $2A$ symbols.

Remark 4. The upgrade-merge-3 procedure replaces three conjugate symbol pairs with two conjugate symbol pairs. Recall from Corollary 21 that after the upgrade-couple transform, conjugate symbols belong to different classes. In particular, if y_a and \bar{y}_a are a conjugate pair of the a-channel before the upgrade-couple transform, then $y_a^{i,j} \in C_{i,j}$ and $\bar{y}_a^{j,i} \in C_{j,i}$ are a conjugate pair of the a-channel after the upgrade-couple transform. Therefore, when one uses Theorem 25 to replace the symbols

$$\{y_{a1}^{i,j}, y_{a2}^{i,j}, y_{a3}^{i,j}\} \rightarrow \{z_{a1}^{i,j}, z_{a3}^{i,j}\},$$

one must also replace their conjugates

$$\{\bar{y}_{a1}^{j,i}, \bar{y}_{a2}^{j,i}, \bar{y}_{a3}^{j,i}\} \rightarrow \{\bar{z}_{a1}^{j,i}, \bar{z}_{a3}^{j,i}\}.$$

We still always operate within a class as nowhere do we mix symbols from different classes. Alternatively, one may upgrade only classes $C_{i,j}$ with $i \geq j$ and then use channel symmetry to obtain the upgraded forms of classes $C_{j,i}$.

There is one case where it is possible to use upgrade-merge-2, as stated in the following corollary.

Corollary 26. Theorem 25 also holds if the a-channel upgrade procedure is upgrade-merge-2 applied to two symbols of the same a-channel D -value.

Proof: While in general the upgrade-merge-2 procedure mixes a symbol and its conjugate, when the two symbols to be merged have the same a-channel D -value this is no longer the case (see Appendix C-A), and we can follow along the lines of the proof of Theorem 25. We omit the details. ■

The reason that [11] introduced both the upgrade-merge-2 and upgrade-merge-3 procedures despite the superiority of the latter stems from numerical issues. To implement upgrade-merge-3 we must divide by the difference of the extremal D -values to be merged. If these are very close this can lead to numerical errors. Upgrade-merge-2 is not susceptible to such errors. On the other hand, upgrade-merge-2 cannot be used in the manner stated above; it requires us to mix symbols from two classes $C_{i,j}$ and $C_{j,i}$ that may have wildly different \tilde{Q}_2 channels. Thus, this will undesirably upgrade the b-channel.

In practice, however, we may be confronted with a triplet of symbols with very close, but not identical, a-channel D -values. To avoid numerical issues, we utilize a fourth nearby symbol.

Say that our triplet⁷ is y_{a1}, y_{a2}, y_{a3} with a-channel D -values $d_{a1} \leq d_{a2} < d_{a3}$ such that $d_{a3} - d_{a1} < \epsilon$, for some “closeness” threshold ϵ . Let y_{a4} have a-channel D -value d_{a4} such that $d_{a4} - d_{a1} > \epsilon$. Then, we apply upgrade-merge-3 twice: first for y_{a1}, y_{a2}, y_{a4} obtaining z_{a1}, z_{a4} with a-channel D -values d_{a1}, d_{a4} and then for z_{a1}, y_{a3}, z_{a4} , ending up with z'_{a1}, z'_{a4} with a-channel D -values d_{a1}, d_{a4} . In this example we have chosen a fourth symbol with a greater a-channel D -value than d_{a4} , but we could have similarly chosen a fourth symbol with a smaller a-channel D -value than d_{a1} instead.

B. Upgrading Channel W_b

We now show how to upgrade $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ to channel $Q_{a,b}(y_a, u_a, z_b|u_a, u_b)$ such that marginal $Q_b \succcurlyeq W_b$ and marginal $Q_a = W_a$. The idea is to begin with W_b^* , a channel equivalent to W_b in which y_a and u_a are not explicit in the output. The channel W_b^* is given by $W_b^*(d_b|u_b) = \sum_{y_a, u_a} W_b(y_a, u_a, d_b|u_b)$. We upgrade W_b^* to Q_b^* using some known method, such that channel P_b^* degrades Q_b^* to W_b^* . To form upgraded channel Q_b , we “split” the outputs of Q_b^* to include y_a and u_a and find a degrading channel that degrades Q_b to W_b . We shall see that the upgraded channel Q_b is given by

$$Q_b(y_a, u_a, z_b|u_b) = Q_b^*(z_b|u_b) \sum_{d_b} \frac{P_b^*(d_b|z_b) W_b(y_a, u_a, d_b)}{W_b^*(d_b)},$$

where $W_b(y_a, u_a, d_b)$ and $W_b^*(d_b)$ are defined in (37), below. Finally, we form the joint channel $Q_{a,b}$ using (8). We illustrate this in Figure 3.

Theorem 27. *Let $W_b(y_a, u_a, d_b|u_b)$ be a joint distribution where d_b is the D -value of the b -channel’s output. Let $W_b^*(d_b|u_b)$ be a channel equivalent to W_b , and let $Q_b^*(z_b|u_b) \succcurlyeq W_b^*(d_b|u_b)$ with degrading channel $P_b^*(d_b|z_b)$. Then there exists joint channel $Q_b(y_a, u_a, z_b|u_b)$ such that $Q_b(y_a, u_a, z_b|u_b) \stackrel{p}{\succcurlyeq} W_b(y_a, u_a, z_b|u_b)$ and $\sum_{y_a, u_a} Q_b(y_a, u_a, z_b|u_b) = Q_b^*(z_b|u_b)$.*

Proof: We shall explicitly find Q_b and an appropriate degrading channel. The degrading channel will be of the form $P_b(d_b|y_a, u_a, z_b)$, i.e., y_a, u_a pass through the degrading channel unchanged. Such degrading channels are proper. Since $Q_b^* \succcurlyeq W_b^*$ we have, for any d_b and u_b ,

$$W_b^*(d_b|u_b) = \sum_{z_b} P_b^*(d_b|z_b) Q_b^*(z_b|u_b). \quad (36)$$

Denote

$$\begin{aligned} W_b(y_a, u_a, d_b) &= \frac{1}{2} \sum_{u_b} W_b(y_a, u_a, d_b|u_b), \\ W_b^*(d_b) &= \frac{1}{2} \sum_{u_b} W_b^*(d_b|u_b). \end{aligned} \quad (37)$$

⁷To simplify notation, we omit the dependence on the class; it is clear that we do this for each class separately.

We assume that $W_b^*(d_b) > 0$, for otherwise output d_b never appears with positive probability and may be ignored, and define

$$\rho_{y_a, u_a}^{d_b} \triangleq \frac{W_b(y_a, u_a, d_b)}{W_b^*(d_b)}.$$

We have $\rho_{y_a, u_a}^{d_b} \geq 0$, $\sum_{y_a, u_a} \rho_{y_a, u_a}^{d_b} = 1$ for any d_b , and, for any u_b ,

$$W_b(y_a, u_a, d_b|u_b) = \rho_{y_a, u_a}^{d_b} W_b^*(d_b|u_b). \quad (38)$$

For each z_b , we will shortly define constants $\mu_{y_a, u_a}^{z_b}$ such that $\mu_{y_a, u_a}^{z_b} \geq 0$ and $\sum_{y_a, u_a} \mu_{y_a, u_a}^{z_b} = 1$. Similar to (38), we use these constants to define channel Q_b by

$$Q_b(y_a, u_a, z_b|u_b) = \mu_{y_a, u_a}^{z_b} Q_b^*(z_b|u_b). \quad (39)$$

Indeed, $\sum_{y_a, u_a} Q_b(y_a, u_a, z_b|u_b) = Q_b^*(z_b|u_b)$. We now find the constants $\mu_{y_a, u_a}^{z_b}$ and an appropriate degrading channel $P_b(d_b|y_a, u_a, z_b)$ such that

$$W_b(y_a, u_a, d_b|u_b) = \sum_{z_b} P_b(d_b|y_a, u_a, z_b) Q_b(y_a, u_a, z_b|u_b), \quad (40)$$

which will establish our goal.

Let y_a, u_a , and d_b be such that the left-hand-side of (40) is positive⁸, so that $\rho_{y_a, u_a}^{d_b} > 0$. We shall see that the resulting expressions hold for the zero case as well. Using (38) and (39), we can rewrite (40) as

$$W_b^*(d_b|u_b) = \sum_{z_b} \left(\frac{P_b(d_b|y_a, u_a, z_b) \mu_{y_a, u_a}^{z_b}}{\rho_{y_a, u_a}^{d_b}} \right) Q_b^*(z_b|u_b).$$

Comparing this with (36), we set

$$P_b^*(d_b|z_b) = \frac{P_b(d_b|y_a, u_a, z_b) \mu_{y_a, u_a}^{z_b}}{\rho_{y_a, u_a}^{d_b}}. \quad (41)$$

Since P_b is a probability distribution, by rearranging and summing over d_b we obtain

$$\mu_{y_a, u_a}^{z_b} = \sum_{d_b} P_b^*(d_b|z_b) \rho_{y_a, u_a}^{d_b}. \quad (42)$$

It is easily verified that $\mu_{y_a, u_a}^{z_b} \geq 0$ and $\sum_{y_a, u_a} \mu_{y_a, u_a}^{z_b} = 1$. Using the expression for $\mu_{y_a, u_a}^{z_b}$ in (41) yields

$$P_b(d_b|y_a, u_a, z_b) = \frac{P_b^*(d_b|z_b) \rho_{y_a, u_a}^{d_b}}{\sum_{d'_b} P_b^*(d'_b|z_b) \rho_{y_a, u_a}^{d'_b}}. \quad (43)$$

This is a valid probability distribution. We remark that (40) is satisfied by (42) and (43) even when $\rho_{y_a, u_a}^{d_b} = 0$. We have found Q_b and a proper degrading channel P_b as required. ■

Corollary 28. *In Theorem 27, the marginal a -channels of Q_b and W_b coincide.*

Proof: By construction, the degrading channel from Q_b to W_b does not change the a -channel output, implying that the a -channel marginal remains the same. ■

⁸Since $W_b^*(d_b) > 0$, there will always be at least one selection of y_a, u_a for which the left-hand-side of (40) is positive.

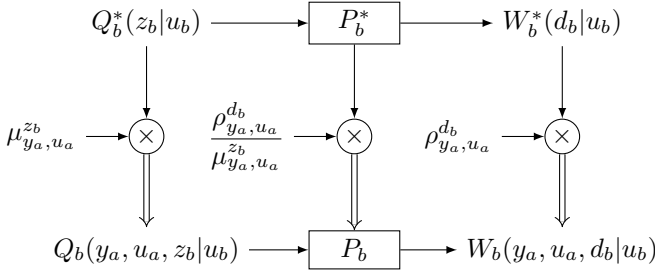


Fig. 3. Illustration of how to transform an upgrading procedure from W_b^* to Q_b^* to an upgrading procedure from W_b to Q_b . The double arrows represent splitting to multiple outputs.

To use Theorem 27, one begins with design parameter B that controls the output alphabet size. The channel Q_b^* , with output alphabet of size $2B$, is obtained from W_b^* using a sequence of upgrade operations. To obtain upgraded joint channel Q_b , one uses the Theorem to turn them into a sequence of upgrade operations to be performed on channel W_b . If one uses the techniques of [11], the upgrade operations will consist of upgrade-merge-2 and upgrade-merge-3 operations (Appendix C). In the following examples we apply Theorem 27 specifically to these upgrades.

For brevity, we will use the following notation:

$$\begin{aligned}\pi_{y_a, u_a}^{d_b} &\triangleq \sum_{u_b} W_b(y_a, u_a, d_b|u_b), \\ \pi^{d_b} &\triangleq \sum_{u_b} W_b^*(d_b|u_b).\end{aligned}\quad (44)$$

Example 4 (Upgrading W_b Based on Upgrade-Merge-2). The upgrade-merge-2 procedure of [11] selects two conjugate symbols pairs and replaces them with a single conjugate symbol pair. The details of the transformation, in our notation, appear in Appendix C-A.

Let joint channel $W_b(y_a, u_a, d_b|u_b)$ have b-channel marginal $W_b^*(d_b|u_b)$, in which all symbols with the same D -value are combined to a single symbol. We select symbols d_{bj}, d_{bk} and their respective conjugates $\bar{d}_{bj} = -d_{bj}, \bar{d}_{bk} = -d_{bk}$, such that $d_{bk} \geq d_{bj} > 0$ and upgrade $W_b^*(d_b|u_b)$ to $Q_b^*(z_b|u_b)$ given by (47) (Appendix C-A). We denote by \mathcal{D}_b the output alphabet of W_b^* and by $\mathcal{D}_{z_{bk}}$ the set

$$\mathcal{D}_{z_{bk}} \triangleq \{d_{bk}, d_{bj}, \bar{d}_{bj}, \bar{d}_{bk}\}.$$

The output alphabet of Q_b^* is $\mathcal{Z} = (\mathcal{D}_b \setminus \mathcal{D}_{z_{bk}}) \cup \{z_{bk}, \bar{z}_{bk}\}$; outputs of Q_b^* represent D -values. In particular, the D -values of z_{bk} and \bar{z}_{bk} are d_{bk} and $-d_{bk}$, respectively.

Using Theorem 27, we form channel $Q_b(y_a, u_a, z_b|u_b)$ by

$$Q_b(y_a, u_a, z_b|u_b) = \begin{cases} \mu_{y_a, u_a}^{z_{bk}} Q_b^*(z_{bk}|u_b) & z_b = z_{bk} \\ \mu_{y_a, u_a}^{\bar{z}_{bk}} Q_b^*(\bar{z}_{bk}|u_b) & z_b = \bar{z}_{bk} \\ W_b(y_a, u_a, z_b|u_b) & \text{otherwise,} \end{cases}$$

where by (42),

$$\mu_{y_a, u_a}^{z_{bk}} = \frac{\sum_{d \in \mathcal{D}_{z_{bk}}} (\pi_{y_a, u_a}^d \cdot (d_{bk} + d))}{2(\pi^{d_{bj}} + \pi^{d_{bk}})d_{bk}},$$

$$\mu_{y_a, u_a}^{\bar{z}_{bk}} = \frac{\sum_{d \in \mathcal{D}_{z_{bk}}} (\pi_{y_a, u_a}^d \cdot (d_{bk} - d))}{2(\pi^{d_{bj}} + \pi^{d_{bk}})d_{bk}}.$$

We can simplify this when W_b is a symmetrized channel. In this case, $\pi_{y_a, u_a}^{d_b} = \pi_{y_a, u_a}^{\bar{d}_b}$, yielding

$$\mu_{y_a, u_a}^{z_{bk}} = \mu_{y_a, u_a}^{\bar{z}_{bk}} = \frac{\pi_{y_a, u_a}^{d_{bj}} + \pi_{y_a, u_a}^{d_{bk}}}{\pi^{d_{bj}} + \pi^{d_{bk}}}.$$

Therefore, the upgraded joint channel becomes

$$Q_b(y_a, u_a, z_b|u_b) = \begin{cases} \Pi_{y_a, u_a}^{z_{bk}} \left(\frac{1 + (-1)^{u_b} d_{bk}}{2} \right) & z_b = z_{bk} \\ \Pi_{y_a, u_a}^{\bar{z}_{bk}} \left(\frac{1 - (-1)^{u_b} d_{bk}}{2} \right) & z_b = \bar{z}_{bk} \\ W_b(y_a, u_a, z_b|u_b) & \text{otherwise,} \end{cases}$$

where

$$\Pi_{y_a, u_a}^{z_{bk}} = (\pi_{y_a, u_a}^{d_{bj}} + \pi_{y_a, u_a}^{d_{bk}}).$$

Example 5 (Upgrading W_b Based on Upgrade-Merge-3). The upgrade-merge-3 procedure replaces three conjugate symbols pairs with two conjugate symbol pairs. The details of the transformation, in our notation, appear in Appendix C-B.

As above, let joint channel $W_b(y_a, u_a, d_b|u_b)$ have b-channel marginal $W_b^*(d_b|u_b)$. For the upgrade procedure we select symbols d_{bi}, d_{bj}, d_{bk} and their respective conjugates, such that $0 \leq d_{bi} < d_{bj} \leq d_{bk}$.⁹ We upgrade $W_b^*(d_b|u_b)$ to $Q_b^*(z_b|u_b)$ given by (48) (Appendix C-A). We denote by \mathcal{D}_b the output alphabet of W_b^* and by $\mathcal{D}_{z_{bk}}$ the set

$$\mathcal{D}_{z_{bk}, z_{bi}} \triangleq \{d_{bk}, d_{bj}, d_{bi}, \bar{d}_{bi}, \bar{d}_{bj}, \bar{d}_{bk}\}.$$

The output alphabet of Q_b^* is $\mathcal{Z} = (\mathcal{D}_b \setminus \mathcal{D}_{z_{bk}, z_{bi}}) \cup \{z_{bk}, z_{bi}, \bar{z}_{bi}, \bar{z}_{bk}\}$; outputs of Q_b^* represent D -values. In particular, the D -values of z_{bk} and z_{bi} are d_{bk} and d_{bi} , respectively.

Assuming that W_b is symmetrized, we form channel $Q_b(y_a, u_a, z_b|u_b)$ using Theorem 27 as

$$Q_b(y_a, u_a, z_b|u_b) = \begin{cases} \mu_{y_a, u_a}^{z_{bk}} Q_b^*(z_{bk}|u_b) & z_b = z_{bk} \\ \mu_{y_a, u_a}^{z_{bi}} Q_b^*(z_{bi}|u_b) & z_b = z_{bi} \\ \mu_{y_a, u_a}^{\bar{z}_{bi}} Q_b^*(\bar{z}_{bi}|u_b) & z_b = \bar{z}_{bi} \\ \mu_{y_a, u_a}^{\bar{z}_{bk}} Q_b^*(\bar{z}_{bk}|u_b) & z_b = \bar{z}_{bk} \\ W_b(y_a, u_a, z_b|u_b) & \text{otherwise,} \end{cases}$$

where by (42),

$$\begin{aligned}\mu_{y_a, u_a}^{z_{bk}} &= \frac{\pi_{y_a, u_a}^{d_{bk}} + \left(\frac{d_{bj} - d_{bi}}{d_{bk} - d_{bi}} \right) \pi_{y_a, u_a}^{d_{bj}}}{\pi^{d_{bk}} + \left(\frac{d_{bj} - d_{bi}}{d_{bk} - d_{bi}} \right) \pi^{d_{bj}}}, \\ \mu_{y_a, u_a}^{z_{bi}} &= \frac{\pi_{y_a, u_a}^{d_{bi}} + \left(\frac{d_{bk} - d_{bj}}{d_{bk} - d_{bi}} \right) \pi_{y_a, u_a}^{d_{bj}}}{\pi^{d_{bi}} + \left(\frac{d_{bk} - d_{bj}}{d_{bk} - d_{bi}} \right) \pi^{d_{bj}}},\end{aligned}$$

and $\mu_{y_a, u_a}^{\bar{z}_{bk}} = \mu_{y_a, u_a}^{z_{bk}}, \mu_{y_a, u_a}^{\bar{z}_{bi}} = \mu_{y_a, u_a}^{z_{bi}}$. The latter two equalities are due to our assumption that W_b is symmetrized.

⁹We could have also selected them such that $0 \leq d_{bi} \leq d_{bj} < d_{bk}$. At least one of the inequalities $d_{bi} \leq d_{bj}$ or $d_{bj} \leq d_{bk}$ must be strict.

Denoting

$$\begin{aligned}\Pi_{y_a, u_a}^{z_{bk}} &= \pi_{y_a, u_a}^{d_{bk}} + \left(\frac{d_{bj} - d_{bi}}{d_{bk} - d_{bi}} \right) \pi_{y_a, u_a}^{d_{bj}} \\ \Pi_{y_a, u_a}^{z_{bi}} &= \pi_{y_a, u_a}^{d_{bi}} + \left(\frac{d_{bk} - d_{bj}}{d_{bk} - d_{bi}} \right) \pi_{y_a, u_a}^{d_{bj}} \\ &= \pi_{y_a, u_a}^{d_{bi}} + \left(1 - \frac{d_{bj} - d_{bi}}{d_{bk} - d_{bi}} \right) \pi_{y_a, u_a}^{d_{bj}},\end{aligned}$$

the upgraded joint channel is given by

$$Q_b(y_a, u_a, z_b | u_b) = \begin{cases} \Pi_{y_a, u_a}^{z_{bk}} \left(\frac{1 + (-1)^{u_b} d_{bk}}{2} \right) & z_b = z_{bk} \\ \Pi_{y_a, u_a}^{z_{bi}} \left(\frac{1 + (-1)^{u_b} d_{bi}}{2} \right) & z_b = z_{bi} \\ \Pi_{y_a, u_a}^{z_{bi}} \left(\frac{1 - (-1)^{u_b} d_{bi}}{2} \right) & z_b = \tilde{z}_{bi} \\ \Pi_{y_a, u_a}^{z_{bk}} \left(\frac{1 - (-1)^{u_b} d_{bk}}{2} \right) & z_b = \tilde{z}_{bk} \\ W_b(y_a, u_a, z_b | u_b) & \text{otherwise.} \end{cases}$$

Remark 5. We observe from these examples an interesting parallel between the a-channel and b-channel upgrading procedures. In the former case, we confine upgrade operations to a single class, in which the b-channel D -values are fixed. In light of the above examples, the latter case may be viewed as confining upgrade procedures to “classes” in which y_a, u_a are fixed.

VII. LOWER BOUND PROCEDURE

The previous sections have introduced several ingredients for building an overall procedure for obtaining a lower bound on the probability of error of polar codes under SC decoding. We now combine these ingredients and present the overall procedure. First, we lower-bound the probability of error of two synthetic channels. Then, we show how to use lower bounds on channel pairs to obtain better lower bounds on the union of many error events.

A. Lower Bound on the Joint Probability of Error of Two Synthetic Channels

We now present an upgrading procedure for $W_{a,b}$ that results in channel $Q_{a,b}$ with a smaller alphabet size. The procedure leverages the recursive nature of polar codes.

The input to our procedure is BMS channel W , the number of polarization steps n , the indices a and b of the a-channel and b-channel, respectively, and parameters A and B that control the output alphabet sizes of the a- and b-channels, respectively. The binary expansions of $a - 1$ and $b - 1$ are $\mathbf{a} = \langle a_1, a_2, \dots, a_m \rangle$ and $\mathbf{b} = \langle b_1, b_2, \dots, b_m \rangle$, respectively. These expansions specify the order of polarization transforms to be performed, where 0 implies a ‘−’-transform and 1 implies a ‘+’-transform.

The algorithm consists of a sequence of polarization and upgrading steps. After each polarization step, we bring the channel to D -value representation, as described in Section IV-A. A side effect of polarization is increase in alphabet size. The

upgrading steps prevents the alphabet size of the channels from growing beyond a predetermined size. After the final upgrading step we obtain joint channel $Q_{a,b}$, which is properly upgraded from $W_{a,b}$. We compute $P_e^{\text{IMJP}}(Q_{a,b})$, which serves as a lower bound to $P_e^{\text{IML}}(W_{a,b})$. We recall that $P_e^{\text{IML}}(W_{a,b})$ is the probability of error under SC decoding of the joint synthetic channel $W_{a,b}$. This, in turn, lower-bounds $P_e^{\text{SC}}(W)$ (see Corollary 17).

Algorithm A provides a high-level description of the procedure. We begin by determining the first index m for which a_m and b_m differ (i.e. $a_\ell = b_\ell$ for $\ell < m$ and $a_m \neq b_m$). The first $m - 1$ polarization steps are of a single channel, as the a-channel and b-channel indices are the same. Since these are single channels, we utilize the upgrading procedures of [11] to reduce the output alphabet size. At the m th polarization step, the a- and b-channels differ. We perform joint polarization described in Section IV-B and symmetrize the channel using (20). This symmetrization need only be performed once as subsequent polarizations maintain symmetrization (Proposition 16). We then perform the b-channel upgrading procedure (Section VI-B), which reduces the b-channel alphabet size to $2B$. Following that, we upgrade the a-channel. As discussed in Section VI-A, this consists of two steps. First, we upgrade-couple the channel, to generate B^2 classes. Second, for each class separately, we use the a-channel upgrade procedure until each class has at most $2A$ elements (see Theorem 25 and Corollary 26). We confine the a-channel upgrade procedure to the class by utilizing only upgrade-merge-3 operations.¹⁰ We continue to polarize and upgrade the joint distribution in this manner, until $\ell = n$. After the final polarization and upgrading operation, we compute the probability of error of the IMJP decoder for the resulting channel.

The lower bound of this procedure compares favorably with the trivial lower bound, $\max\{\mathbb{P}\{\mathcal{E}_a\}, \mathbb{P}\{\mathcal{E}_b\}\}$. This is because our upgrading procedure only ever changes one marginal, keeping the other intact. Since it leverages upgrading transforms that can be used on single channels, the marginal channels obtained are the same as would be obtained on single channels using the same upgrading steps. Thus, by Lemma 7 this lower bound is at least as good as $\max\{\mathbb{P}\{\mathcal{E}_a\}, \mathbb{P}\{\mathcal{E}_b\}\}$.

Remark 6. When the BMS W is a BEC, we can recover the bounds of [8] and [10] using our upgrading procedure. Only a-channel upgrades are required, as the b-channel, in D -value representation, remains a BEC. For each a-channel symbol, the channel W_2 in (22) is either a perfect channel or a pure-noise channel (see Lemma 31 in Appendix A). Thus, the upgrade-couple procedure splits the a-channel symbols to those that see a perfect channel regardless of u_a and those that see a pure-noise channel regardless of u_a . Merging a-channel symbols of the same class is equivalent to merging a-channel symbols for which \tilde{W}_2 is the same type of channel. We thus merge a-channel symbols of the same a-channel D -value that “see” the same type of b-channel. This corresponds to keeping track of the correlation between erasure events of the two channels.

¹⁰When merging symbols of the same a-channel D -value we may also use the upgrade-merge-2 procedure.

Algorithm A: A lower bound on the probability of error under SC decoding of a joint synthetic channel

Input: BMS channel W , number of polarization steps n , channel indices a, b , and alphabet-size control parameters A, B . The binary representations of $a - 1$ and $b - 1$ are $\mathbf{a} = \langle a_1, a_2, \dots, a_n \rangle$ and $\mathbf{b} = \langle b_1, b_2, \dots, b_n \rangle$, respectively.

Output: A lower bound on the probability of error $W_{a,b}$.

```

 $m \leftarrow \text{first\_difference}(\mathbf{a}, \mathbf{b})$ 
 $Q \leftarrow \text{single\_upgrade}(W, \max\{A, B\})$ 
for  $\ell = 1, 2, \dots, n$  do
  if  $\ell < m$  then
     $Q \leftarrow \text{single\_polarize}(Q, a_\ell)$ 
     $Q \leftarrow \text{D-Value\_representation}(Q)$ 
     $Q \leftarrow \text{single\_upgrade}(Q, \max\{A, B\})$ 
  else
     $Q \leftarrow \text{jointly\_polarize}(Q, a_\ell, b_\ell)$ 
     $Q \leftarrow \text{D-Value\_representation}(Q)$ 
    if  $\ell = m$  then
       $Q \leftarrow \text{symmetrize}(Q)$ 
      // b-channel upgrade:
       $Q \leftarrow \text{b-channel\_upgrade}(Q, B)$ 
      // a-channel upgrade:
       $Q \leftarrow \text{upgrade\_couple}(Q)$ 
      foreach  $\text{class} \in Q$  do
         $Q \leftarrow \text{a-channel\_upgrade}(Q, A, \text{class})$ 
        /* Confine to class by using
           only upgrade-merge-3. */
    return  $P_e^{\text{IMJP}}(Q)$ 

```

Remark 7. An initial step of Algorithm A is to upgrade the channel W , even before any polarization operations. This step enables us to apply our algorithm on continuous-output channels, see [11, Section VI].

B. Lower Bound for More than Two Synthetic channels

Recall that the probability of error of polar codes under SC decoding may be expressed as $\mathbb{P}\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\}$. In the previous section, we developed a lower bound on $\mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}$, which lower bounds $\mathbb{P}\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\}$. This lower bound may be strengthened by considering several pairs of synthetic channels and using (3). We now show how this can be done.

Lemma 29. *The probability of error of a union of events $\bigcup_{i \in \mathcal{A}} \mathcal{E}_i$ is lower bounded by*

$$\mathbb{P}\left\{\bigcup_{i \in \mathcal{A}} \mathcal{E}_i\right\} \geq \sum_{\substack{i, j \in \mathcal{A}, \\ i < j}} \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\} - (|\mathcal{A}| - 2) \sum_{i \in \mathcal{A}} \mathbb{P}\{\mathcal{E}_i\}.$$

Proof: The proof hinges on using the identity $\mathbb{P}\{\mathcal{E}_i \cap \mathcal{E}_j\} = \mathbb{P}\{\mathcal{E}_i\} + \mathbb{P}\{\mathcal{E}_j\} - \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}$ in (3). Note

that any set of $|\mathcal{A}|$ numbers $\{p_i\}$ satisfies

$$\begin{aligned} 2|\mathcal{A}| \sum_i p_i &= \sum_{i, j} (p_i + p_j) \\ &= \sum_{i < j} (p_i + p_j) + \sum_{i = j} (p_i + p_j) + \sum_{i > j} (p_i + p_j) \\ &= 2 \sum_{i < j} (p_i + p_j) + 2 \sum_i p_i, \end{aligned}$$

so that

$$\sum_{i < j} (p_i + p_j) = (|\mathcal{A}| - 1) \sum_i p_i.$$

Therefore,

$$\begin{aligned} \sum_{i < j} \mathbb{P}\{\mathcal{E}_i \cap \mathcal{E}_j\} &= \sum_{i < j} (\mathbb{P}\{\mathcal{E}_i\} + \mathbb{P}\{\mathcal{E}_j\} - \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}) \\ &= (|\mathcal{A}| - 1) \sum_i \mathbb{P}\{\mathcal{E}_i\} - \sum_{i < j} \mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}. \end{aligned}$$

Using this in (3) yields the desired bound. \blacksquare

In practice, we combine the lower bound of Lemma 29 with (13). I.e., we compute lower bounds on $\mathbb{P}\{\mathcal{E}_i \cup \mathcal{E}_j\}$ for all pairs of channels in some subset \mathcal{A}' of the non-frozen set, and use Lemma 29 over this subset.

Such bounds are highly dependent on the selection of the subset \mathcal{A}' . One possible strategy is as follows. Let \mathcal{B} be the set of k worst channels in the non-frozen set for some k . For each channel pair in \mathcal{B} , compute a lower bound on the joint probability of error using Algorithm A. Then, form all possible subsets of \mathcal{B} (there are 2^k such subsets) and use Lemma 29 for each subset. Choose the subset with the highest upper bound as \mathcal{A}' . The reason for going over all possible subsets is that bounds based on the inclusion-exclusion principle are not guaranteed to be higher than the highest pairwise probability, see [14].

VIII. NUMERICAL RESULTS

Bounds on the probability of error of a polar code of length $N = 2^{10}$ over a BSC are shown in Figure 4. We designed the polar code for a BSC with crossover probability 0.2 using the techniques of [11] with 128 quantization levels. As the non-frozen set \mathcal{A} we selected the 102 channels with smallest probability of error, to yield a code rate of approximately 0.1. This non-frozen set was fixed. Then, for BSCs with various crossover probabilities, we computed the following bounds. For the upper bound, we computed an upper bound on $\sum_{a \in \mathcal{A}} P_e(W_a)$, and for the trivial lower bound we computed a lower bound on $\max_{a \in \mathcal{A}} P_e(W_a)$; upper and lower bounds on $P_e(W_a)$ were obtained using the techniques of [11]. For the pair and combination lower bounds we used lower bounds on the IMJP decoder, described in this paper. These were computed with $2A = 32$ and $2B = 8$ for all possible pairs of the 10 worst channels in the non-frozen set. The pair lower bound is merely the highest probability of error of all pairs, whereas the combination lower bound is based on Lemma 29 computed for the subset of these 10 channels that yielded the highest bound.

As one may observe, our bounds improve upon the previously known lower bound (2). Further, the combination lower bound is, as may be expected, tighter than the pair lower bound.

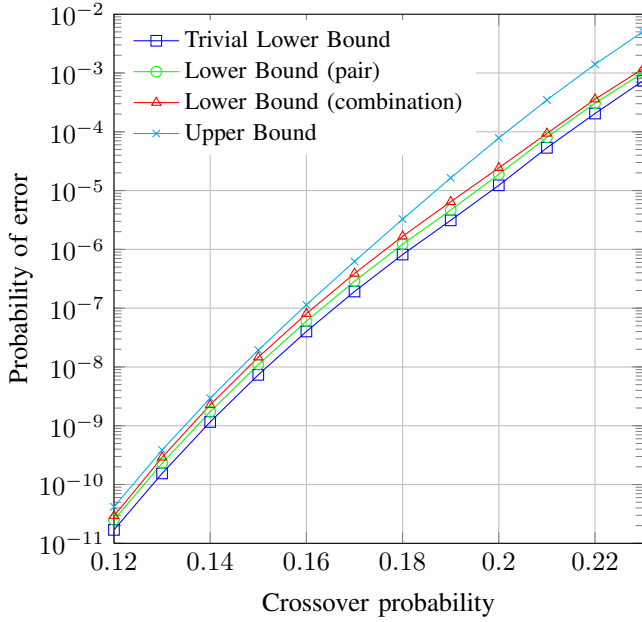


Fig. 4. Bounds on the probability of error of a rate $1/10$ polar code of length 2^{10} designed for a BSC with crossover probability 0.2. The code was used over BSCs with a range of crossover probabilities. The upper bound is based on [11]. The trivial lower bound is a lower bound on $\max_{a \in \mathcal{A}} P_e(W_a)$. The worst pair and worst combination lower bounds were computed using the techniques of this paper.

APPENDIX A THE IMJP DECODER FOR A BEC

In the special case where W is a BEC and W_a and W_b are two of its polar descendants, we have the following.

Proposition 30. *Let $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ be two polar descendants of a BEC in the same tier. Then, the IMJP and the IML (SC) decoders coincide.*

To prove this, we first show that for the BEC erasures are determined by the received channel symbols, $y_1^{2^n}$, and not previous bit decisions. This implies that for fixed y_a , regardless of y_r and in particular u_a , either channel W_b always experiences an erasure, or always experiences a non-erasure. If W_b experiences an erasure, it doesn't matter what ϕ_a decides in terms of the IMJP decoder – it may as well use an ML decoder; if W_b does not experience an erasure, then the best bet of W_a is to use an ML decoder. This suggests that the IML and IMJP decoders coincide.

Lemma 31. *Let $W_i(y_1^{2^n}, u_1^{i-1}|u_i)$ be a polar descendant of a BEC, W . Then, there exists a set K_n , dependent only on i , such that W_i has an erasure if and only if $y_1^{2^n} \in K_n$.*

Proof: Here, $y_1^{2^n}$ are the received channel symbols, and u_1^{i-1} the previous bit decisions that are part of W_i 's output. Let $\langle b_1, b_2, \dots, b_n \rangle$ be the binary expansion of $i-1$, with b_1 the MSB. Recall that channel W_i is the result of n polarization steps determined by b_1, b_2, \dots, b_n , where $b_j = 0$ is a ‘-’-transform and $b_j = 1$ is a ‘+’-transform.

Consider first the case where $n = 1$, i.e., $i-1 = b_1$. If $b_1 = 0$ then $W_i = W^-$ has an erasure if and only if at

least one of y_1, y_2 is an erasure, i.e., if and only if $y_1^2 \in K_1$, $K_1 = \{y_1^2 | y_1 = e \text{ or } y_2 = e\}$. If $b_1 = 1$ then $W_i = W^+$ has an erasure if and only if both y_1 and y_2 are erasures, i.e., if and only if $y_1^2 \in K_1$, $K_1 = \{y_1^2 | y_1 = e \text{ and } y_2 = e\}$. Therefore, the claim is true for $n = 1$.

We proceed by induction. Let the claim be true for $n-1$: for $i'-1 = \langle b_1, b_2, \dots, b_{n-1} \rangle$, there exists a set K_{n-1} such that $W_{i'}$ has an erasure if and only if $y_1^{2^{n-1}} \in K_{n-1}$. If $b_n = 0$, then W_i is the result of a ‘-’-transform of two BEC channels $W_{i'}$, so it has an erasure if and only if at least one of them erases. In other words, W_i has an erasure if and only if $y_1^{2^n} \in K_n$, $K_n = \{y_1^{2^n} | y_1^{2^{n-1}} \in K_{n-1} \text{ or } y_{2^{n-1}+1}^{2^{n-1}} \in K_{n-1}\}$. If, however, $b_n = 1$, then W_i is the result of a ‘+’-transform of two BEC channels $W_{i'}$, so it has an erasure if and only if both of them erase. In other words, W_i has an erasure if and only if $y_1^{2^n} \in K_n$, $K_n = \{y_1^{2^n} | y_1^{2^{n-1}} \in K_{n-1} \text{ and } y_{2^{n-1}+1}^{2^{n-1}} \in K_{n-1}\}$. Thus, the claim is true for n as well, completing the proof. ■

Proof of Proposition 30: By Lemma 3, a decoder ϕ_b that minimizes $\mathbb{P}\{\mathcal{E}_a \cup \mathcal{E}_b\}$ is an ML decoder. It remains to show that a minimizing ϕ_a is also an ML decoder. Marginalizing the joint distribution (8) yields W_a :

$$W_a(y_a|u_a) = \sum_{u_b, y_b} W_b(y_b|u_b) \mathbb{I}[y_b = (y_a, u_a, y_r)].$$

The ML decoder for channel W_a maximizes $W_a(y_a|u_a)$ with respect to u_a ; decoder ϕ_a , on the other hand, maximizes $T(y_a|u_a)$, defined in (10). Using (8) we recast the expression for T in the same form as the expression for W_a ,

$$\begin{aligned} T(y_a|u_a) &= \sum_{u_b, y_b} W_b(y_b|u_b) \mathbb{I}[y_b = (y_a, u_a, y_r)] \cdot \mathbb{P}\{\phi_b(y_b) = u_b\}. \end{aligned}$$

By Lemma 31, whether W_b has an erasure depends solely on the received channel symbols, which are wholly contained in y_a , and not on previous bit decisions. In particular, in computing W_a or T , we either sum over only erasure symbols or over only non-erasure symbols. Since ϕ_b is an ML decoder for W_b , if y_b is an erasure of W_b then $W_a(y_a|u_a) = 2T(y_a|u_a)$; if y_b is not an erasure of W_b then $W_a(y_a|u_a) = T(y_a|u_a)$. In either case, it is clear that the decision based on (9) is identical to the ML decision. Therefore, ϕ_a is an ML decoder as well, implying that the IMJP decoder is an IML decoder. ■

APPENDIX B INTRODUCTION TO D -VALUES

The decision of an ML decoder for a memoryless binary-input channel $W_{Y|U}$ may be based on any sufficient statistic of the channel output. One well-known sufficient statistic is the log-likelihood ratio (LLR), $l(y) = \log \left(\frac{W_{Y|U}(y|0)}{W_{Y|U}(y|1)} \right)$. When $l(y)$ is positive, the decoder declares that 0 was transmitted; when $l(y)$ is negative, the decoder declares that 1 was transmitted; $l(y) = 0$ constitutes an erasure, at which the decoder makes some random choice. Another sufficient statistic is the D -value.

The D -value of output y , $d(y)$, is given by

$$d(y) \triangleq W_{U|Y}(0|y) - W_{U|Y}(1|y). \quad (45)$$

Clearly, $-1 \leq d(y) \leq 1$. A maximum likelihood decoder makes its decision based on the sign of the D -value. Assuming a symmetric channel input, $U = 0, 1$ with probability $1/2$, using Bayes' law on (45) yields

$$d(y) = \frac{W_{Y|U}(y|0) - W_{Y|U}(y|1)}{W_{Y|U}(y|0) + W_{Y|U}(y|1)} \quad (46)$$

The input is binary, hence $W_{U|Y}(0|y) + W_{U|Y}(1|y) = 1$. Consequently (46) yields

$$\begin{aligned} \frac{1 + d(y)}{2} &= \frac{W_{Y|U}(y|0)}{W_{Y|U}(y|0) + W_{Y|U}(y|1)} = W_{U|Y}(0|y), \\ \frac{1 - d(y)}{2} &= \frac{W_{Y|U}(y|1)}{W_{Y|U}(y|0) + W_{Y|U}(y|1)} = W_{U|Y}(1|y). \end{aligned}$$

There is a one-to-one correspondence between $d(y)$ and $l(y)$, $l(y) = \log \frac{1+d(y)}{1-d(y)}$, or, equivalently, $d(y) = \tanh(l(y)/2)$.

If channel $W_{Y|U}$ is symmetric, for each output y there is a conjugate output \bar{y} ; their LLRs and D -values are related: $l(\bar{y}) = \frac{1}{l(y)}$, $d(\bar{y}) = -d(y)$.

Since the D -value is a sufficient statistic of a BMS channel, we may replace the channel output with its D -value. Thus, we may assume that the output y of channel $W_{Y|U}$ is a D -value, i.e., $y = W_{U|Y}(0|y) - W_{U|Y}(1|y)$. In this case, we say that W is in D -value representation.

Recall that every BMS channel can be decomposed into BSCs [15, Theorem 2.1]. We can think of the output of a BMS as consisting of the “reliability” of the BSC and its output. The absolute value of the D -value corresponds to the BSC's reliability and its sign to the BSC output (0 or 1).

A comprehensive treatment of D -values and LLRs in relation to BMS channels appears in [12, Chapter 4].

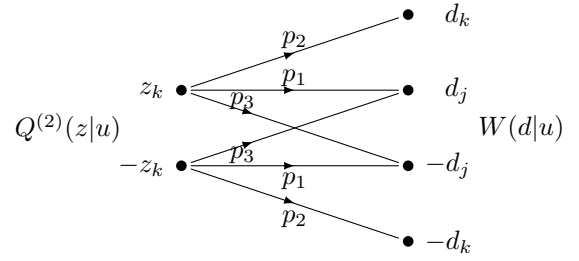
APPENDIX C UPGRADES OF A BMS CHANNEL

We state here in our notation the two upgrades of a BMS channel from [11].

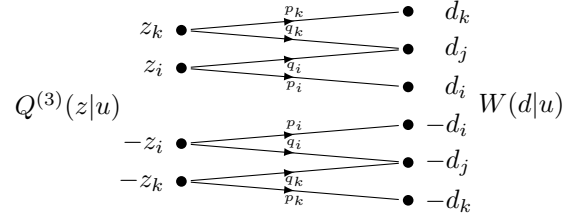
Let W be a discrete BMS whose outputs are D -values $\pm d_1, \pm d_2, \dots, \pm d_m$, and let the probability of symbol d_ℓ be $\pi^{d_\ell} \triangleq W(d_\ell|u) + W(-d_\ell|u) = W(d_\ell|0) + W(d_\ell|1)$, $\ell = 1, \dots, m$. Without loss of generality, $0 \leq d_1 \leq d_2 \leq \dots \leq d_m \leq 1$. Clearly, $\pi^{d_\ell} \geq 0$ for all ℓ , and $\sum_{\ell=1}^m \pi^{d_\ell} = 1$. Moreover, $\pi^{d_\ell} = \pi^{-d_\ell}$. I.e., this is a BMS that decomposes to m different BSCs, with crossover probabilities $(1 - d_\ell)/2$, $\ell = 1, \dots, m$. BSC channel ℓ is selected with probability π^{d_ℓ} . We have $W(d_\ell|u) = (\pi^{d_\ell}/2) \cdot (1 + (-1)^u d_\ell)$ and $W(-d_\ell|u) = W(d_\ell|\bar{u})$.

A. The Upgrade-merge-2 Procedure

The first upgrade-merge of [11] takes two D -values $d_j \leq d_k$ and merges them by transferring the probability of d_j to d_k . We call it *upgrade-merge-2*. Channel $W : \mathcal{U} \rightarrow \mathcal{Y}$ is upgraded



(a) Degrading channel from $Q^{(2)}$ to W for upgrade-merge-2.



(b) Degrading channel from $Q^{(3)}$ to W for upgrade-merge-3.

Fig. 5. Degrading channels for the upgrade-merge-2 and upgrade-merge-3 procedures.

to channel $Q^{(2)} : \mathcal{U} \rightarrow \mathcal{Z}$; the output alphabet of $Q^{(2)}$ is $\mathcal{Z} = (\mathcal{Y} \setminus \{d_j, d_k, -d_j, -d_k\}) \cup \{z_k, -z_k\}$, and

$$Q^{(2)}(z|u) = \begin{cases} \pi^{z_k} \left(\frac{1 + (-1)^u d_k}{2} \right) & z = z_k \\ \pi^{z_k} \left(\frac{1 - (-1)^u d_k}{2} \right) & z = -z_k \\ W(z|u) & \text{otherwise,} \end{cases} \quad (47)$$

where

$$\pi^{z_\ell} = \begin{cases} 0 & \ell = j \\ \pi^{d_j} + \pi^{d_k} & \ell = k \\ \pi^{d_\ell} & \text{otherwise.} \end{cases}$$

The degrading channel from $Q^{(2)}$ to W is shown in Figure 5a. We show only the portion of interest, i.e., we do not show the symbols that this degrading channel does not change. The parameters of the degrading channel are

$$\begin{aligned} p_1 &= \frac{\pi^{d_j}}{\pi^{d_j} + \pi^{d_k}} \left(\frac{d_k + d_j}{2d_k} \right), \\ p_2 &= \frac{\pi^{d_k}}{\pi^{d_j} + \pi^{d_k}}, \\ p_3 &= \frac{\pi^{d_j}}{\pi^{d_j} + \pi^{d_k}} \left(\frac{d_k - d_j}{2d_k} \right). \end{aligned}$$

Indeed, $p_1, p_2, p_3 \geq 0$ and $p_1 + p_2 + p_3 = 1$, so this constitutes a valid channel. Note that if $d_j = d_k$ then $p_3 = 0$.

B. The Upgrade-merge-3 Procedure

The second upgrade-merge of [11] removes a D -value d_j by splitting its probability between a preceding D -value $d_i \leq d_j$ and a succeeding D -value $d_k \geq d_j$. We call it *upgrade-merge-3*. Unlike upgrade-merge-2, at least one of these inequalities must be strict (i.e., either $d_i < d_j$ or $d_j < d_k$). Channel $W : \mathcal{U} \rightarrow \mathcal{Y}$

is upgraded to channel $Q^{(3)} : \mathcal{U} \rightarrow \mathcal{Z}$ with output alphabet $\mathcal{Z} = (\mathcal{Y} \setminus \{d_i, d_j, d_k, -d_i, -d_j, -d_k\}) \cup \{z_i, z_k, -z_i, -z_k\}$, and

$$Q^{(3)}(z|u) = \begin{cases} \pi^{z_k} \left(\frac{1 + (-1)^u d_k}{2} \right) & z = z_k \\ \pi^{z_i} \left(\frac{1 + (-1)^u d_i}{2} \right) & z = z_i \\ \pi^{z_i} \left(\frac{1 - (-1)^u d_i}{2} \right) & z = \bar{z}_i \\ \pi^{z_k} \left(\frac{1 - (-1)^u d_k}{2} \right) & z = \bar{z}_k \\ W(z|u) & \text{otherwise,} \end{cases} \quad (48)$$

where

$$\pi^{z_\ell} = \begin{cases} \pi^{d_i} + \pi^{d_j} \left(\frac{d_k - d_j}{d_k - d_i} \right) & \ell = i \\ 0 & \ell = j \\ \pi^{d_k} + \pi^{d_j} \left(\frac{d_j - d_i}{d_k - d_i} \right) & \ell = k \\ \pi_\ell & \text{otherwise.} \end{cases}$$

Note that

$$Q^{(3)}(z_k|u) + Q^{(3)}(z_i|u) = W(d_i|u) + W(d_j|u) + W(d_k|u). \quad (49)$$

The degrading channel from $Q^{(3)}(z|u)$ to $W(y|u)$ is shown in Figure 5b, showing only the interesting portion of the channel. The parameters of the channel are $p_\ell = \pi^{d_\ell} / \pi^{z_\ell}$, and $q_\ell = 1 - p_\ell$, $\ell = i, k$. This is a valid channel as $\pi^{z_\ell} \geq \pi^{d_\ell}$.

It can be shown [11, Lemma 12] that $Q^{(2)} \succcurlyeq Q^{(3)} \succcurlyeq W$. I.e., upgrade-merge-3 yields a better (closer) upgraded approximation of W than does upgrade-merge-2.

REFERENCES

- [1] E. Arkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [2] —, "Source polarization," in *2010 IEEE International Symposium on Information Theory*, June 2010, pp. 899–903.
- [3] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2009.
- [4] E. Şaşıoğlu, "Polarization and polar codes," *Foundations and Trends in Communications and Information Theory*, vol. 8, no. 4, pp. 259–381, 2011.
- [5] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, December 2013.
- [6] E. Şaşıoğlu, "Polarization in the presence of memory," in *2011 IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 189–193.
- [7] E. Şaşıoğlu and I. Tal, "Polar coding for processes with memory," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 225–229.
- [8] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE International Symposium on Information Theory*, June 2009, pp. 1496–1500.
- [9] F. M. Hoppe, "Iterating Bonferroni bounds," *Statistics & Probability Letters*, vol. 3, no. 3, pp. 121–125, 1985.
- [10] M. Bastani Parizi and E. Telatar, "On the correlation between polarized BECs," in *Proc. IEEE International Symposium on Information Theory*, July 2013, pp. 784–788.

- [11] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.
- [12] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.
- [13] D. E. Knuth, "Two notes on notation," *Am. Math. Monthly*, vol. 99, no. 5, pp. 403–422, May 1992.
- [14] S. J. Schwager, "Bonferroni sometimes loses," *The American Statistician*, vol. 38, no. 3, pp. 192–197, 1984.
- [15] I. Land and J. Huber, "Information combining," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 3, pp. 227–330, 2006.